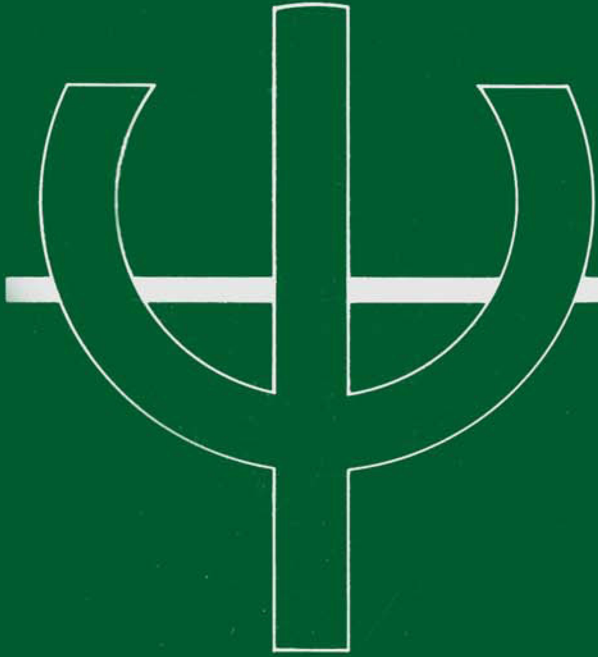


ADVANCES
IN
PSYCHOLOGY

20

Trends in Mathematical Psychology

Edited by
E. Degreef
J. van Buggenhaut

North-Holland

SHOUT
SHARE
IT

TRENDS IN MATHEMATICAL PSYCHOLOGY

ADVANCES IN PSYCHOLOGY

20

Editors

G. E. STELMACH

P. A. VROON

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

TRENDS IN
MATHEMATICAL PSYCHOLOGY

Edited by

E. DEGREEF
and
J. VAN BUGGENHAUT
Centre for Statistics
and Operational Research
Brussels, Belgium

1984

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD

© Elsevier Science Publishers B.V., 1984

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording or otherwise, without the prior
permission of the copyright owner.

ISBN: 0 444 87512 3

Publishers:
ELSEVIER SCIENCE PUBLISHERS B.V.
P.O. Box 1991
1000 BZ Amsterdam
The Netherlands

Sole distributors for the U.S.A. and Canada:
ELSEVIER SCIENCE PUBLISHING COMPANY, INC.
52 Vanderbilt Avenue
New York, N.Y. 10017
U.S.A.

Library of Congress Cataloging in Publication Data
Main entry under title:

Trends in mathematical psychology.

(Advances in psychology ; 20)
Papers presented at the 14th European Mathematical Psychology Group Meeting, held in Brussels, Sept. 12-14, 1983.
Includes indexes.
Psychometrics--Congresses. I. Degreef, E., 1953- II. Buggenhaut, J. van, 1937- . III. European Mathematical Psychology Group. IV. European Mathematical Psychology Group. Meeting (14th : 1983 : Brussels, Belgium) V. Series: Advances in psychology (Amsterdam, Netherlands) ; 20.

BF39.T74 1984 150'.28'7 84-6032
ISBN 0-444-87512-3 (U.S.)

PRINTED IN THE NETHERLANDS

PREFACE

This volume gathers most of the papers presented at the 14th European Mathematical Psychology Group Meeting. The meeting took place in Brussels from September 12 to September 14, 1983 and was hosted by the Vrije Universiteit Brussel. Financial support was also provided by the Belgian "Nationaal Fonds voor Wetenschappelijk Onderzoek", the Belgian "Ministerie van Nationale Opvoeding", the Belgian "Culturele Betrekkingen" and the "Nationale Bank van België".

The idea of the Group is to bring together once a year, people in Europe working in the field of Mathematical Psychology. Taken into account the growing importance attached to mathematical models in human sciences, the experiences acquired in previous meetings and the presence, besides other well-known participants, of four invited lecturers, namely S. Grossberg (Boston University), F.S. Roberts (Rutgers, State University of New Jersey) and V.F. Venda - V.Y. Krylov (Moscow, USSR Academy of Sciences), gave us the idea to make a selection of the contributions and to gather them in a book.

In order to structure the whole, we took the liberty to group the papers into three parts, knowing that the classification can be discussed; some of the papers, indeed, can find a place in more than one part. So you will find:

Part I : Perception, learning and memory
Part II : Order and measurement
Part III: Data analysis.

We hope that the studies collected here, fairly represent the different perspectives and that the volume as a whole will be a dynamic resource for those who want to keep abreast of Mathematical Psychology in general and the European one in particular.

We finally wish to thank Francis Gheys for the excellent typing of the manuscript. It was a lot of work, but we enjoyed it.

The editors
Brussels, February 1984

This Page Intentionally Left Blank

CONTENTS

Preface v

Participants xi

PART I: PERCEPTION, LEARNING AND MEMORY 1

Tree representations of associative structures in semantic and episodic memory research
H. Abdi, J.-P. Barthélémy, X. Luong 3

Task-dependent representation of categories and memory-guided inference during classification
H. Buffart, H.-G. Geissler 33

Outline of a theory of brightness, color, and form perception
S. Grossberg 59

Attention in detection theory
H.C. Micko 87

Impossible objects and inconsistent interpretations
E. Térouanne 105

In searching of general regularities of adaptation dynamics: on the transformation learning theory
V.F. Venda 121

A probabilistic choice model adapted for the angle perception experiment
N.D. Verhelst 159

PART II: ORDER AND MEASUREMENT 175

About the asymmetries of similarity judgments: an ordinal point of view
J.-P. Barthélémy 177

The axiomatization of additive difference models for preference judgments
M.A. Croon 193

Generalizations of interval orders
J.-P. Doignon 209

Réseaux sociaux généralisés: en combinant graphes et hypergraphes
C. Flament 219

On concepts of the dimension of a relation and a generalized recognition experiment
K. Herbst 227

Isotonic regression analysis and additivity
R.P. Mac Donald 239

Testing Fechnerian scalability by maximum likelihood estimation of ordered binomial parameters
R. Mausfeld 255

Probabilistic consistency, homogeneous families of relations and linear Λ-relations
B. Monjardet 271

Applications of the theory of meaningfulness to order and matching experiments
F.S. Roberts 283

A new derivation of the Rasch model
E.E. Roskam, P.G.W. Jansen 293

A definition of partial interval orders
M. Roubens, Ph. Vincke 309

Causal linear stochastic dependencies: the formal theory
R. Steyer 317

Factorization and additive decomposition of a weak order
R. Suck 347

PART III: DATA ANALYSIS 363

The problem of representation based upon two criteria
G. De Meur, M. Gassner, X. Hubaut 365

Tree representations of rectangular proximity matrices
G. De Soete, W.S. DeSarbo, G.W. Furnas, J.D. Carroll 377

Weak and strong models in order to detect and measure poverty
P. Dickes 393

Applications of a Bayesian Poisson model for misreadings
M.G.H. Jansen 405

The polychotomous Rasch model and dichotomization of graded responses
P.G.W. Jansen, E.E. Roskam 413

Component analysis of transition probabilities and its application to prisoner's dilemma
K. Rainio 433

An application of multidimensional scaling on prediction: the Radex structure
F. Van Overwalle 449

Author index 473

Subject index 477

This Page Intentionally Left Blank

PARTICIPANTS

H. ABDI, Laboratoire de psychologie, Ancienne Faculté, Rue Chabot Charny, 21000 Dijon, France
H. ALGAYER, Donndorforstr. 93, D-8580 Bayreuth, Federal Republic Germany
J. ANDRES, Rathausstr. 53, 53 Bonn 3, Federal Republic Germany
J.-P. BARTHELEMY, ENST, Département d'Informatique, 46 Rue Barrault 75634 Paris Cedex 13, France
A. BOHRER, CRS, Sectie voor Psychologisch Onderzoek, Kazerne Klein Kasteeltje, 1000 Brussel, Belgium
H.F.J.M. BUFFART, Psychologisch Laboratorium, Katholieke Universiteit Nijmegen, Postbus 9104, 6500HE Nijmegen, The Netherlands
M.A. CROON, Katholieke Hogeschool Tilburg, Hogeschoollaan 225, Tilburg, The Netherlands
E. DEGREEF, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
L. DELBEKE, Psychologisch Instituut, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven, Belgium
G. DE MEUR, Université Libre de Bruxelles, CP 135, 1050 Bruxelles, Belgium
G. DE SOETE, Dienst voor Psychologie, Rijksuniversiteit Gent, Henri Dunantlaan 2, 9000 Gent, Belgium
M. DESPONTIN, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
P. DICKES, Université de Nancy II, Bd Albert I, BP 3397, 54015 Nancy-Cedex, France
J.-P. DOIGNON, Université Libre de Bruxelles, CP 216, 1050 Bruxelles, Belgium
C. FLAMENT, Département de Psychologie, Université de Provence, Les Blaques, Cereste, F-04110 Reillanne, France
H.-G. GEISSLER, Sektion Psychologie, Karl-Marx-Universität Leipzig, Tieckstr. 2, 7030 Leipzig, German Democratic Republic
M. GASSNER, Université Libre de Bruxelles, CP 135, 1050 Bruxelles, Belgium
C.A.W. GLAS, C.I.T.O., Postbus 1034, 6801 MG Arnhem, The Netherlands
S. GROSSBERG, Center for Adaptive Systems, Boston University, 111 Cummington Street, Boston, Massachusetts 02215, USA
P.K.G. GUNTHER, Siebengebirgstr. 11, D-5330 Konigswinter 41, Federal Republic Germany
M. HAHN, Mollwitzstr. 5, D-1000 Berlin 19, Federal Republic Germany
K. HERBST, Institut für Psychologie, Universität Regensburg, Universitätstr. 31, D-8400 Regensburg, Federal Republic Germany
D. HEYER, Institut für Psychologie, Universität Kiel, Ohlshausenstr. 40/60, D-2300 Kiel, Federal Republic Germany
M.G.H. JANSEN, Instituut voor Onderwijskunde, Rijksuniversiteit Groningen, Westerhaven 16, 9718 AW Groningen, The Netherlands
V.Y. KRYLOV, Department of Mathematical Psychology, Institute of Psychology, USSR, Academy of Sciences, Moscow 129366, 13 Yaroslavskaya, USSR
N. LAKENBRINK, Auf dem Kampf 11, D-2000 Hamburg 63, Federal Republic Germany
X. LUONG, Laboratoire de Mathématiques, Université de Besançon, Besançon, France
R.R. MAC DONALD, Department of Psychology, University of Stirling, Stirling FK94LA, Scotland

A.W. MAC RAE, University of Birmingham, Box 363, Birmingham B15 2TT, Great Britain
R. MAUSFELD, Universität Bonn, Auf den Steinen 13A, 53 Bonn 1, Federal Republic Germany
H.C. MICKO, Institut für Psychologie, Technische Universität Braunschweig, Spielmannstrasse 12A, D-3300 Braunschweig, Federal Republic Germany
B. MONJARDET, Centre de Mathématique Sociale, 54 Bd Raspail, 75270 Paris Cedex 06, France
C. MULLER, Lehrstuhl für Psychologie III, Universität Regensburg, Postfach 397, D-8400 Regensburg, Federal Republic Germany
P. PLENNIGER, Seminar für Philosophie und Errichungswissenschaft der Universität Freiburg, Werthmannplatz, D-7800 Freiburg, Federal Republic Germany
B. ORTH, Institut für Psychologie, Universität Hamburg, Von-Melle-Park 6, D-2000 Hamburg 13, Federal Republic Germany
M. PIRLOT, Université de l'Etat à Mons, 17 Place Warocqué, 7000 Mons, Belgium
J.G.W. RAAIJMAKERS, Psychologisch Laboratorium, Katholieke Universiteit Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands
K. RAINIO, University of Helsinki, Petaksentie 44, 00630 Helsinki, Finland
F.S. ROBERTS, Rutgers, State University of New Jersey, Dep. of Math. at New Brunswick, Hill Center for the Mathematical Sciences, Bush Campus, New Brunswick, New Jersey 08903, U.S.A.
E.E. ROSKAM, Vakgroep Mathematische Psychologie, Katholieke Universiteit Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands
M. ROUBENS, Faculté Polytechnique de Mons, 9 Rue de Houdain, B-7000 Mons, Belgium
U. SCHULZ, Fakultät für Psychologie und Sportwissenschaft, Abt. für experimentelle und angew. Psychologie, Universität Bielefeld, Postfach 8640, 4800 Bielefeld, Federal Republic Germany
A.J. SMOLENAARS, Psychologisch Laboratorium, Universiteit Amsterdam, Weesperplein 8, 1018 XA Amsterdam, The Netherlands
R. STEYER, Fachbereich I - Psychologie, Universität Trier, Schneidershof, D-5500 Trier, Federal Republic Germany
R. SUCK, Universität Osnabrück, Postfach 4469, 45 Osnabrück, Federal Republic Germany
E. TEROUANNE, UER Mathématiques, Université Paul Valéry, BP 5043, 34032 Montpellier - Cedex, France
P. VAN ACKER, 1 Chaussée de Wavre, 1050 Bruxelles, Belgium
J. VAN BUGGENHAUT, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
A. VAN DER WILDT, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
A. VAN DINGENEN, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
L. VAN LANGENHOVE, Faculteit Psychologie en Opvoedkunde, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
F. VAN OVERWALLE, EDUCO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
J.C. VAN SNICK, Faculté des Sciences Economiques et Sociales, Université de l'Etat à Mons, 17 Place Warocqué, 7000 Mons, Belgium
V.F. VENDA, Department of Engineering Psychology, Institute of Psychology, USSR Academy of Sciences, Moscow 129366, 13 Yaroslavskaya, USSR
R. VERHAERT, CSOO, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium

N.D. VERHELST, Subfaculteit der psychologie, vakgroep PSM, Rijksuniversiteit Utrecht, Sint-Jacobsstraat 14, 3511 BS Utrecht, The Netherlands
Ph. VINCKE, Institut de Statistique, Université Libre de Bruxelles, CP 210, B-1050 Bruxelles, Belgium

This Page Intentionally Left Blank

PART I

PERCEPTION, LEARNING AND MEMORY

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

TREE REPRESENTATIONS OF ASSOCIATIVE STRUCTURES IN SEMANTIC AND EPISODIC MEMORY RESEARCH

Hervé Abdi
Laboratoire de Psychologie, Dijon

Jean-Pierre Barthélémy
E.N.S.T., Paris

Xuan Luong
Laboratoire de Mathématiques, Besançon

We expose some research in the area of psychology of memory involving proximity or distance matrices. We propose some ways of building up such matrices. Then we detail an algorithm allowing the representation of proximity matrices by an additive tree, and contrast this new algorithm with previous ones. Finally, we examine some results obtained with this method.

1. INTRODUCTION

The general purpose of this paper is to emphasize the utility and describe the use of additive trees in order to describe data collected in the field of the psychology of memory. This paper is threefold: we first describe some research leading to the construction of distance or proximity matrices; secondly we expose and detail the construction of an additive tree as a representation of the original matrix; finally, we examine the results obtained.

The utilization of clustering methods for attesting the organization of memory or revealing its structure has been strongly advocated recently by some authors in different areas of cognitive psychology (see, among others: Miller (1969), Henley (1969), Friendly (1978), Rosenberg et al (1968), (1972), (1982)). Most of the used methods amount to represent the original matrix by an Ultrametric Tree. Recently, there has been an attempt to build some methods leading to representations less stringent than the classical Ultrametric Tree, i.e. the Additive Tree (see Carroll & Chang (1973), Cunningham (1974), (1978); Sattath & Tversky (1977)). We propose hereafter an (economical) heuristic giving an Additive Tree from a proximity matrix and illustrate it with some examples borrowing from our current research or from classical papers.

2. A BUNCH OF EXAMPLES

2.1. BARTLETT'S "THE WAR OF THE GHOSTS"

In 1932, Bartlett asked a few subjects to read an American Indian folk tale (named "The War of the Ghosts") and to recall the story on several occasions (a method called "repeated reproduction"); in a variant of the method (i.e. "serial reproduction") a chain of different subjects is used, the first being shown the original text and then recalling for the second subject who would pass it to a third and so on. These classical experiments of Bartlett will serve here to illustrate a set of informatic procedures, the aim of which is to build some distances between texts.

A) Informatic procedures

For reasons of compatibility, the programs are written in Standard Microsoft Basic (Under CP/M), and at least 64 K-Bytes of RAM and a disk Unit are needed. Although these programs include some various possibilities, we will restrict ourself to the part dealing specifically with the construction of metrics between texts.

It must be clear that when we speak of the text given by a subject, we could speak as well of a set of themes or ideas given by a subject providing an adequate coding of the raw data.

The texts are first transformed in a disk file, then for each text we build the Lexicon associated with it. This Lexicon could be either a Boolean Lexicon (i.e. it merely indicates the Presence or the Absence of the item of Vocabulary) or an integer Lexicon (i.e. it indicates the number of Occurrences of each item). From the different Lexicons (Boolean or Integer) we build - by union - a General Lexicon that defines the Vocabulary shared by the different texts.

B) Construction of distances between texts

Depending on the point of view adopted, we could define different distances; as an illustration we examine three ways:

(i) the texts as subsets of the Vocabulary
(ii) the texts as Bi-partitions of the Vocabulary
(iii)a "probabilistic"generalization.

Denote by Li the Lexicon associated with a text Ti, the general Lexicon by $V = \bigcup_{i} Li$, and by $\overline{Li}$ the complement of Li in V.

(i) Each (Boolean) Lexicon* is a subset of the Vocabulary and we could use, for example, the well-known distance between sets, the so called cardinal of the symmetric difference:

$$d(Ti,Tj) = |Li \,\Delta\, Lj| = |Li \cap \overline{Lj}| + |\overline{Li} \cap Lj|.$$

(ii) $\{Li,\overline{Li}\}$ defines a Bi-Partition of V (i.e. a Partition with two classes), and so does $\{Lj,\overline{Lj}\}$. So, we could use some distances between Partitions (cf. Arabie & Boorman (1973)) or Bi-Partitions,e.g. the distance of the symmetric difference between Bi-partitions that can be expressed as:

$$\begin{aligned} d(Ti,Tj) &= 2(|Li\cap Lj|+|\overline{Li}\cap\overline{Lj}|)(|Li\cap\overline{Lj}|+|\overline{Li}\cap Lj|) \\ &= 2(|\overline{Li\Delta Lj}|)(|Li\Delta Lj|) \end{aligned}$$

(iii) In order to take explicitly account of the Integer Lexicons we could look for an extension of (i). With each Ti is associated a probability measure on V (i.e. the frequency of the different words); denote the probability of item x of text Ti by Pi(x); then we find a family of distances by

$$d_q(Ti,Tj) = \sum_{x\in V} [Pi(x)-Pj(x)]^q$$

2.2. FEATURES OF PERSONALITY

This research lays on the border between the work on the organization of the semantic memory and the work upon the "implicit psychology". The purpose is to describe the subjective organization of the qualifiers of the character. As a matter of fact, it has often been noted that we tend to group subjectively some features of character as if we has an "Implicit Theory of Personality" (cf. e.g., Rosenberg et al (1972), Wegner and Vallacher (1977)). In this experiment we select fifty eight qualifiers of the character (using some Thesauruses and a bit of literature,...). These qualifiers are then printed on separate cards and given individually to twenty-eight subjects with the request that he or she sort the cards into piles with the constraint that "the cards in a same pile give the feeling to go together"; subjects were free to choose the number of piles for sorting (for a review of the pro and contra of the sorting method, see Rosenberg (1982)).

---

*Notice that one could " fuzzy " the "boolean" distance in (i) and (ii) by taking the fuzzy equivalent of the union and intersection, i.e. Min and Max.

Hence, each subject expresses his opinion by a partition on the set of the qualifiers, and for using as a Q-Methodology (cf. Kerlinger (1973)) the afore evoked distances between partitions could easily be used. The partition given by a subject is associated with a matrix whose rows and columns represent the qualifiers, and where we put a 1 at the intersection of a row and a column if the qualifiers are not sorted in the same pile by the subject. Obviously this is a distance matrix (cf. Miller (1969)), and so will be the matrix defined by the sum of the matrices of the different subjects. In this matrix we simply count the number of subjects who do not put together the qualifiers. Dually we could have defined a matrix of co-occurences by the sum of the so-called incidence matrix (where a 1 means that the qualifiers are in the same pile), for commodity reasons this is the matrix we give later (cf. Table 3).

It must be noted, in passing, that the Data obtained and consequently the distance matrix depend upon the methods designed for obtaining such Data. In particular, other methods (e.g. word associations, or distances between words in free-recall, etc.) lead to other results (see Abdi (1983)).

2.3. OLDIES BUT GOLDIES

It could be useful to compare different methods on well-known data. We will examine here three sets of data (included two "classics").

The first comes from a study by Miller (1969) pioneering the structure of semantic memory. Fifty college students, each sorted forty-eight words according to their similarity of meaning (the subjects were given a brief definition of each noun taken from a dictionary). The co-occurence matrix (similar to the matrix previously discussed, cf. 2.2) was subjected to hierarchical clustering using Johnson's (1967) connectedness and diameter method. These results are recalled in 4 (cf. Table 4 and Figure 5).

The second is extracted from a study conducted by Henley (1969). She obtained from twenty one subjects an estimation of the distance between twelve animal terms (for each subject she simply counts the number of terms separating the terms of interest in a list given by the subject provided with the instruction to "list all the animals they could"); the matrices are then standardized and the mean for each cell (e.g. across subjects) give the entry of a dissimilarity matrix.

Finally, Friendly (1979), in a paper akind to the two previous ones, strongly advocate the use of hierarchical clustering to diagram the memory organization. He proposed a proximity matrix obtained somewhat like Henley except that the original data came from free recall experiments. The data were "a priori" organized in three subsets: animal terms, parts of the human body, vegetables. The aim of the study was the recovery of this a priori structure (cf. Bousfield (1953), Tulving (1972)).

3. METHODS

Broadly speaking, the methods dealing with proximity data can be divided under two general categories: The Maps and the Graphs.

The Maps encompass - in particular - the different variants of Multidimensional Scaling (cf. Kruskall (1978), Schiffman et al (1981), Young et al, to appear), Optimal or Dual Scaling or Correspondence Analysis depending on the authors (cf. Benzecri (1973), Hill (1974), Nishisato (1980)) or classical scaling or "triple analysis" if the proximity could be judged as distance (Torgerson (1958), Benzecri(1973)).

The general characteristic ofthe Graphs is to represent the objects as vertices and the proximity (or the distance) between them by arcs (with or without valuation). One could possibly explore the structure of these graphs by looking for the connected components, or its centroid (cf. Abdi (1980)) or by imposing some threshold for drawing the arcs between vertices (cf. The Analysis of Similitude developed by Flament et al (1962), and Degenne and Verges (1973)). One can eventually impose some restrictions on the graph, i.e. being a tree. We detail this point in a moment.

3.1. TREES

A general definition of a Tree can be a cycle free, connected, undirected graph; for convenience assume a valuation on the edges, and define the distance between two vertices as the length of the path from one vertex to the other.

The Ultrametric Tree could be characterized - besides other conditions - by the classical " ultrametric inequality". Three vertices - say x,y,z - on an Ultrametric Tree verify:

$$d(x,y) \leqslant \text{Max}\ [d(x,z)\ ,\ d(y,z)]\ .$$

This property is quite drastic and barely verified by similarity measures obtained from an experiment (see Miller (1969) for a stimulating discussion).

So the problem is to build up the better ultrametric approximation of a given similarity matrix, and a lot of algorithms are available for this job (cf. Sneath and Sokal (1973), Hartigan (1971)). Instead of imposing the ultrametric inequality on the representation of a dissimilarity (or distance) matrix, some authors have proposed to weaken the ultrametric inequality in order to obtain a more general and more natural representation (cf. Carrol and Chang (1973), Cunningham (1974), (1978); Sattath and Tversky (1977) called following the authors weighted tree, free tree, path length tree or unrooted tree (cf. among others: Riordan (1958), Hakami and Yau (1964), Buneman (1971), Dobson (1974)). This kind of a tree is characterized by the following inequality, holding for every four-uplet - say x,y,u,v -:

$$d(x,y) + d(u,v) \leq \text{Max } [d(x,u) + d(y,v);\ d(x,v) + d(y,u)] \ .$$

It can be shown that the ultrametric inequality is stronger than the additive inequality which, in turn, implies the triangle inequality.

The algorithms for fitting an additive tree to a proximity matrix are less numerous than those designed for the species ultrametric tree. The construction of the additive tree can be separated (for the clarity of the explanation) into two parts:

1) the finding of the tree-structure (we dare not use the term skeleton...)
2) the estimation of the valuation of the branches of the tree.

The classical approach (as illustrated by Cunningham (1974), (1978); Sattath and Tversky (1977)) makes a direct use of the additive inequality for the finding of the tree-structure, and then estimates the valuation with a least-square method. For our part, we propose another approach that we detail in a moment and contrast with the classical one.

3.2. CONSTRUCTION OF THE TREE STRUCTURE

(A) Sattath and Tversky (1977) introduced - in a seminar paper - the notion of "loose cluster".

Let E be a set and d a dissimilarity function defined on E x E, hereafter denoted as $\langle E,d \rangle$. A subset A from E is a loose cluster if, for all $\{x,y\}$ from A and for all $\{u,v\}$ from E (with $x \neq u$, $x \neq v$, $y \neq u$, $y \neq v$), the following inequality is verified.

$$d(x,y) + d(u,v) < \text{Min } [d(x,u) + d(y,v);\ d(x,v) + d(y,u)] \ . \quad (1)$$

If E corresponds to the terminal vertices of a tree, then the following inequality holds for all four-uplet (x,y,u,v):

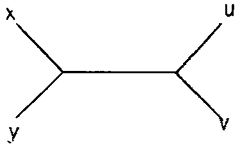
$$d(x,y) + d(u,v) \leq d(x,u) + d(y,v) \leq d(x,v) + d(y,u) \qquad (2)$$

(or any of the five inequalities that can be deduced from this one by permutation).

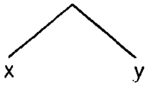
Then if {x,y} is a loose cluster on a tree, from (2) and (1) follows:

$$d(x,y) + d(u,v) < d(x,u) + d(y,v) = d(x,v) + d(y,u) \ .$$

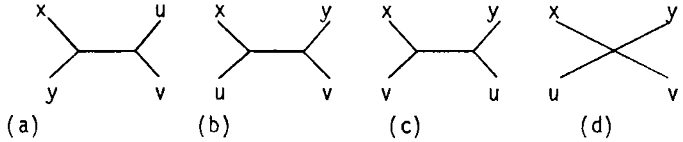
So, in the following tree with terminal vertices x,y,u,v



{x,y} on the one hand, and {u,v} on the other are loose clusters, and (2) is then defined. In particular, x and y, u and v are to adopt the following configuration and are said to merge.



Conversely, knowing that x,y,u,v could be represented by a tree, we get only four possible cases:



Here, the knowledge of a loose cluster is sufficient to characterize fully the representation.

(i) If at least one pair is loose then:

{x,y} loose ⇔ {u,v} loose ⇔ tree a
{x,u} loose ⇔ {y,v} loose ⇔ tree b
{x,v} loose ⇔ {y,u} loose ⇔ tree c

(ii) No pair is loose, and then we get the tree d.

These previous remarks could be used in order to build the tree structure fitting a ⟨E,d⟩.

For each {x,y} from E the Sattath and Tversky's algorithm counts the number of pairs $\{u,v\}$ verifying (1); call this number the score of the pair $\{x,y\}$. The algorithm then merges the highest score pair - say x,y - into a new element - say z -. The dissimilarity between z and the remaining elements is simply obtained by the average of the dissimilarity of x,y. The process is then reiterated untill all elements have been merged. It can easily be verified that this algorithm gives the exact tree of $\langle E,d \rangle$ if it exists, providing that the interior vertices of the tree are of degree 3.

(B) We first give some definitions.

(i) Scoring subset and score.
We use the same notations as before. Let k be an integer less than the cardinality of E. $\{a,b\}$ is element of $E \times E$, and $E^* = E - \{a,b\}$. Then we say that

a k-subset - say S - of E^* is a scoring subset for $\{a,b\}$ if for all x,y from S the following inequality holds:

$$d(a,b) + \mathrm{Max}\ d(x,y) < \mathrm{Max}[d(a,x) + d(,by)] \ .$$

The number of k-scoring subsets for the pair $\{a,b\}$ is called the k-score of $\{a,b\}$ and is denoted $k\text{-}Sc(a,b)$.

Let $\langle E,d \rangle$ be an additive tree and k equal to 2. The so-called canonical distance gives to every branch of the tree the length 1. Now we can give the:

Proposition: If the canonical distance between two terminal vertices - say a,b - is two, then $k\text{-}Sc(a,b)$ is maximal and $k\text{-}Sc(a,u) = k\text{-}Sc(b,u)$ for every u of E.

The proof is left to the reader, but can eventually be found in Luong (1983) with some other properties of the scores.

(ii) Construction of the tree structure.
Set $k=2$. For every pair x,y we count the number of scoring subsets. We compute then a score matrix. We look for the pair with the maximal score - say a,b -, then we merge a,b to give c and we pose:

$$k\text{-}Sc(c,u) = [k\text{-}Sc(a,u) + k\text{-}Sc(b,u)] \ / \ 2 \ .$$

The process is then reiterated until all the remaining elements have been merged. This procedure is justified by the previous proposition. Notice in passing that we could straightforwardly extend this procedure to any k (cf. Luong (1983)).

3.3. ESTIMATION OF THE VALUATION OF THE EDGES

(A) Sattath and Tversky (1977), Cunningham (1974), (1978), propose a least square method for estimating the length of the branches of the tree. Precisely, if d are the distances between terminal vertices on the tree and δ the dissimilarity observed, d is to satisfy:

$$\mathrm{MIN}\left[\sum_{x,y\in E}[d(x,y)-\delta(x,y)]^2\right]$$

This problem is equivalent to the matricial equation:

$$C^t\, C\, X = C^t\, \delta,$$

where δ is the vector of the observed dissimilarity, C the incidence matrix of the branches and t denotes the transposed matrix.

Moreover Sattath and Tversky indicate the existence - but without giving it - of a method avoiding the computation of the inverse of a matrix.

(B) The general idea of our procedure is to make a geometrical estimation of the distance; i.e. we embed the dissimilarities in an Euclidean space, and use the geometrical properties of this space in order to obtain an estimation of the distances on the tree.

The tree obtained as described above, is a binary tree. If d is a distance defined on a binary tree, representing the pair {a,b} by z, the best possible representation puts z on the nearest vertex from a,b. In this case, the distances d(z,u) - u belonging to E - are easily estimated:

$$d(z,u) = [d(a,u) + d(b,u) - d(a,b)]\ /\ 2\ .$$

In the more general case where d is a dissimilarity, we introduce a "central point" called g, with:

$$d(g,u) = \left[\sum_{v\in E} d(v,u)\right] / n,$$

for all u in E, with n = cardinality of E.

We then embed the dissimilarities in an Euclidean space. If {a,b} has a maximal score, we build the subtree a z b. The distances on the tree will then be evaluated by the remaining dissimilarities d(a,u) d(b,u) for all u from E. Now z represents {a,b}; we make the hypothesis that z must be brought nearer to g. According to this, we can determine by the geometry the new dissimilarity d(z,u) for all u from E - {a,b}. The processus is then reiterated parallelly with the score procedure described previously.

3.4. EVALUATION OF THE ALGORITHM

The procedure we proposed is clearly heuristic, and so, it could be interesting to have an evaluation of the quality of it. One possible way can be to generate some random distance matrices, to build the additive tree, then to rebuild the distance matrix and to measure the degree of fit between the original matrix and the rebuilt matrix. And so we did. We decided to choose as a measure of fit the classical product moment correlation, but it is well known that r is strongly related to some other measures of fit, e.g. the stress of Kruskall. The results of our simulations are given in table 1.

Number of trials	set size	average r	standard deviation	Min	Max
25	12	.886	.036	.84	.95
20	18	.884	.034	.833	.930
16	24	.861	.029	.811	.914
18	30	.863	.016	.843	.88
10	36	.860	.022	.82	.89
8	42	.855	.018	.81	.905
4	48	.840	.021	.82	.875

Table 1

Goodness of fit measure between random Euclidean distance matrices and rebuilt distance on an additive tree.

It could be seen that the overall degree of fit is quite good. In fact, the results obtained fall closely near the results obtained by Pruzansky et al (1982) with the original ADDTREE of Tversky.

3.5. AVAILABILITY

Our algorithm is translated into standard Basic (CP/M) and a listing of the program is available on request.

4. SOME RESULTS

The different figures and tables are given at the end of the paper.

4.1. THE WAR OF THE GHOSTS

We examine here the results obtained with the distance between texts seen as Bi-partitions (as defined in 2.1.B.ii). Figure 1 gives the additive tree obtained with the "repeated reproduction" method. We have included the original text for ease of comparison. For convenience, we give the distance matrix (Table 1) and the Map resulting of a Factorial Analysis of Distance (F.A.D.) applied to this matrix (Figure 2).

One striking result is the fidelity of the subjects to themselves as could be seen on the additive tree. Moreover, one could easily detect the accurate subject(s) here, for example, the subject LP is fairly accurate both on the same day or 120 days after. These results are worth noting, because the emphasis in the literature is generally put on the Gist of the recall for these kind of data instead of Verbatim recall. Nevertheless, the fidelity of the subjects to themselves could be due to the stability of their habits of language as well as to the "quality" or constancy of their memory. The FAD Map gives essentially the same conclusion, athough the dimensions revealed by the Map are not easily interpreted. Nevertheless a discrepancy must be noted: the Map indicates L 120 far from ORIG and L1 constrating with the additive tree which gives L 120 near ORIG and L1. In fact the examination of the data matrix reveals that L 120 and L1 are close to each other, so do L1 and ORIG but ORIG and L 120 are quite far from each other. So the two methods differ when a general view is not obvious.

The additive tree obtained from the distance matrix for the "serial reproduction" method allows the recovery of the serial structure (i.e. a semi-order) of the data (although the additive tree model is probably not the better one for this kind of problem). As previously, a simple look at the tree suffices to distinguish the inaccurate "Gaps" (e.g. between R3 and R4, and R6 and R7).

4.2. THE QUALIFIERS OF THE CHARACTER

The data matrix 58 x 58 represents actually our biggest example treated by our methods. The efficiency of our algorithm appears here: we needed seven hours (on a micro computer) to obtain the tree and the valuations on the arcs; in contrast, the classical algorithm would have needed -at the very least - three days in the same conditions.

We have included here in addition to the additive tree, the three-dimensional result of a F.A.D. of the same data. Although there is no major discrepancy between the two approaches, the additive tree is clearly more readable. Probably because the implicit model here is one of cluster rather than one of different dimensions. To be more precise, the sorting instruction induces the construction of clusters and the high degree of agreement between subjects (cf. the data matrix) leads obviously to a "cluster" solution. Nevertheless, some distinct clusters of qualifiers can be identified and be interpreted in terms of "implicit psychology", but this interpretation remains to be verified experimentally, later on.

4.3. OLDIES BUT GOLDIES

For the different examples evoked, we give the original data matrix, the hierarchical solution provided by the authors and the additive tree (except for Henley for which we give a Map resulting from a F.A.D.). The comparison is left to the reader as a passtime....

TABLES AND FIGURES

	ORIG	N1	N15	HP1	HP8	L1	L120	P1	P45	P105	P1145	R1	R15	R45	X180
ORIG	0														
N1	81	0													
N15	81	32	0												
HP1	119	102	106	0											
HP8	121	102	108	42	0										
L1	78	85	89	107	113	0									
L120	110	91	99	97	99	88	0								
P1	118	83	101	91	97	102	92	0							
P45	109	88	94	100	102	107	95	59	0						
P105	113	92	104	94	100	105	97	49	52	0					
P1145	115	80	84	76	80	107	87	71	80	76	0				
R1	80	59	65	75	83	92	82	80	77	87	53	0			
R15	91	62	68	76	84	93	79	75	78	86	46	23	0		
R45	101	68	68	78	92	103	77	79	86	90	42	37	20	0	
X180	113	78	90	84	94	111	87	83	92	90	58	65	58	56	0

Table 2

Distance matrix between 15 texts obtained by Bartlett (1932) with the method of "repeated reproduction".

The distance is the distance of the symmetrical difference between Bi-partitions (see text). The subjects are denoted by the letter(s) beginning the Label, the figures following the letter(s) indicate the number of days between the first presentation of "the war of the Ghosts" and the recall.

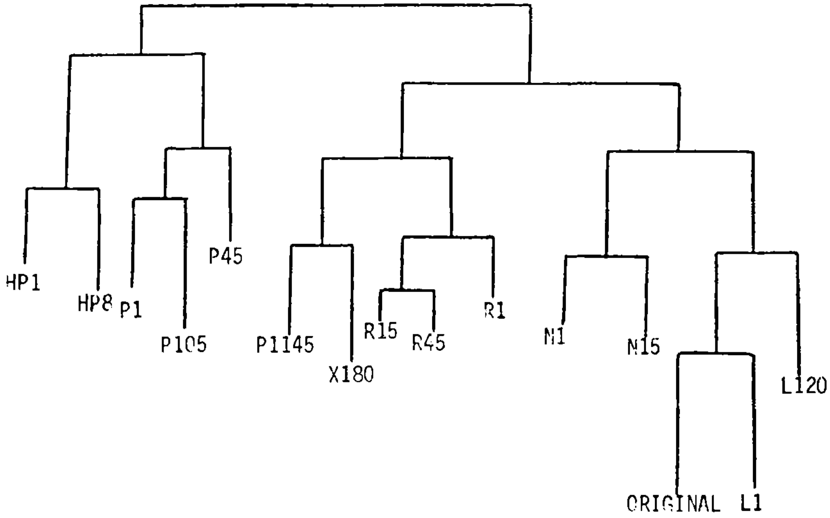

HP1
HP8
P1
P105
P45
P1145
X180
R15
R45
R1
N1
N15
ORIGINAL
L1
L120

Figure 1

Additive tree obtained from the distance matrix of Table 2. The subjects are denoted by the letter(s) beginning the label, the figure following the letter(s) indicate the number of days between the first presentation of the "War of the Ghosts" and the recall.

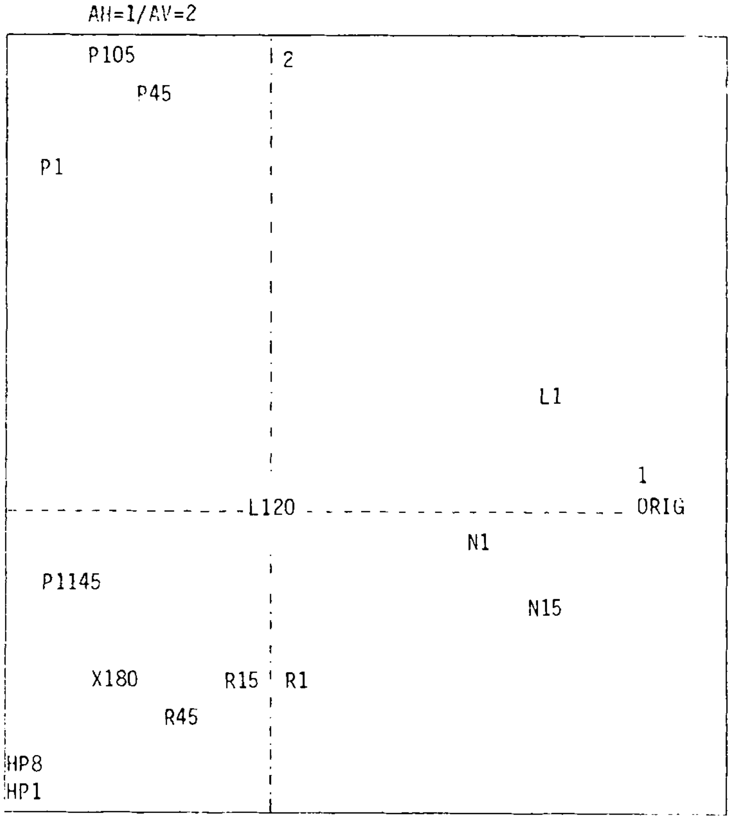


Figure 2

Euclidean representation (factorial analysis of distance) of the distance matrix of Table 2.

The subjects are denoted by the letter(s) beginning the label, the figures following the letter(s) indicate the number of days between the first presentation of "the War of the Ghosts" and the recall.

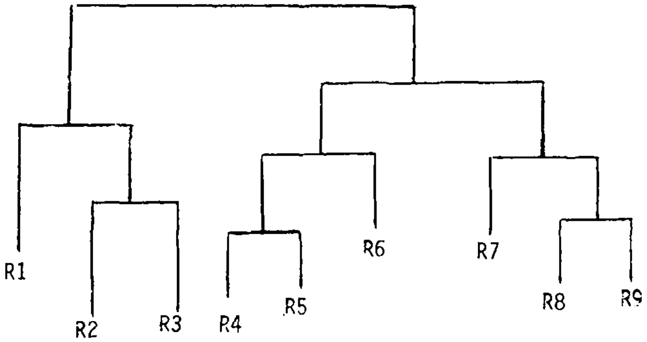


Figure 3

Additive tree obtained from the distance matrix built from the data of Bartlett (1932): "Serial reproduction".
The labels give the serial order.

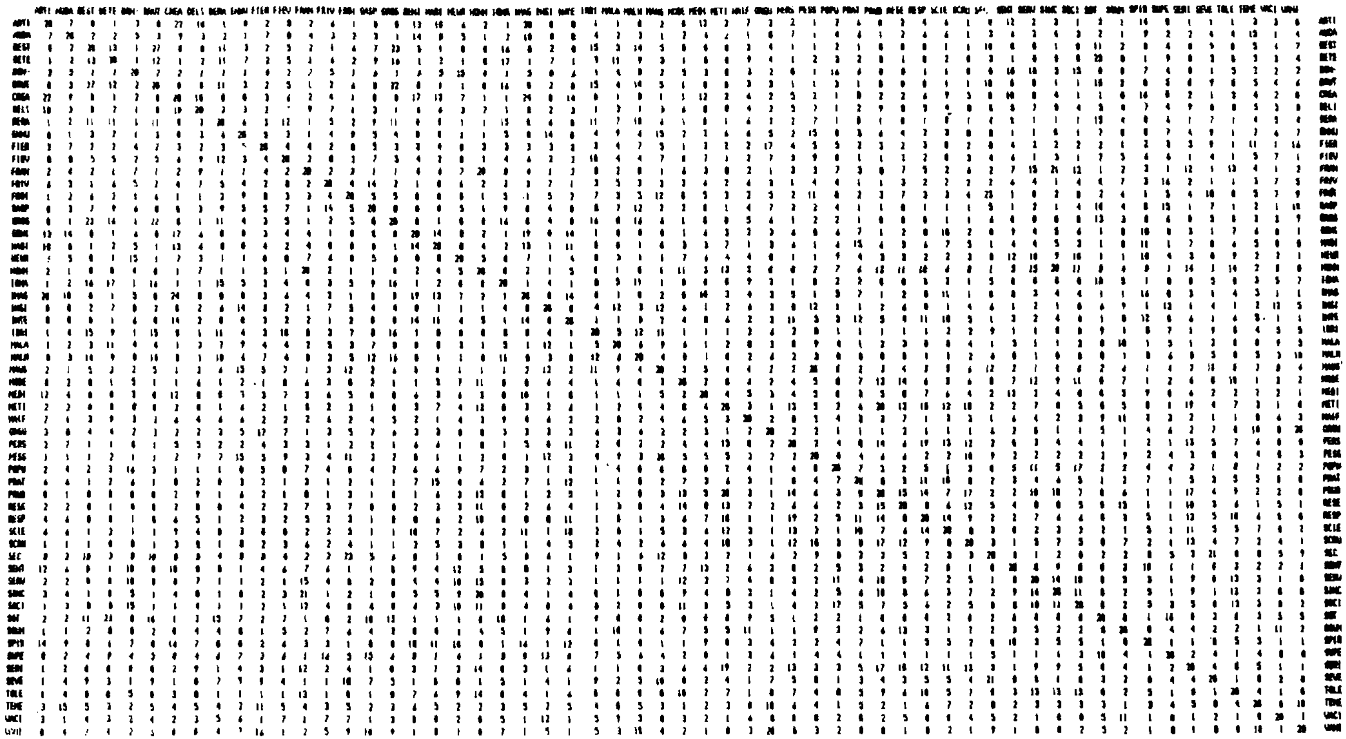


Table 3

Co-occurence Matrix for the Qualifiers of the character. The cells indicate the number of Subjects who put the qualifiers in the sample Pile (number of subjects: 28). The labels indicate the first four letters of the words.

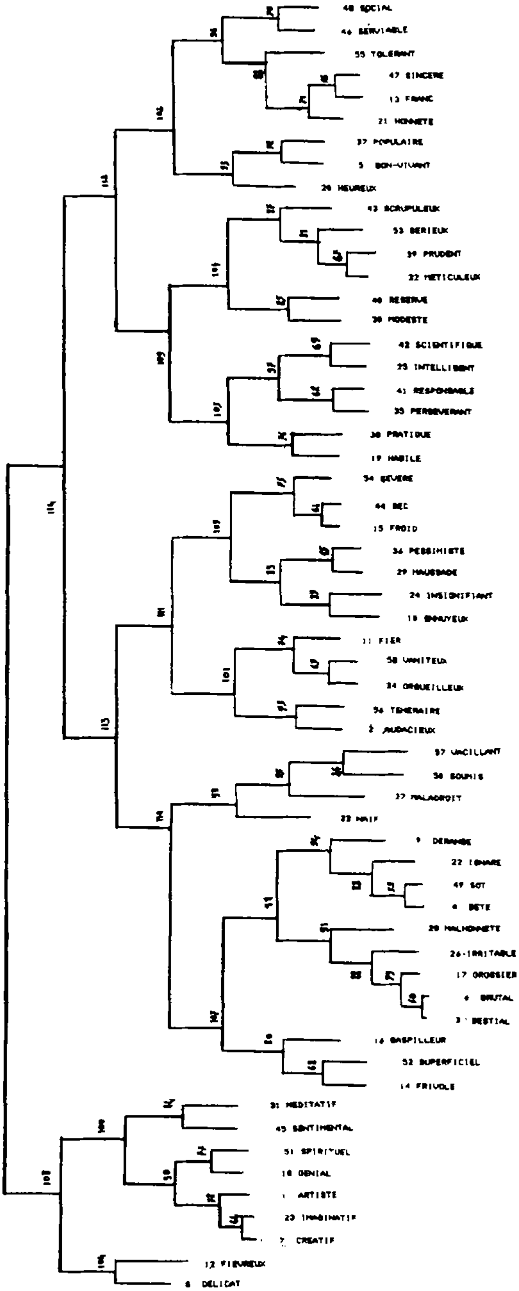

48 SOCIAL
46 SERVIABLE
55 TOLERANT
47 SINCERE
13 FRANC
21 HONNETE
37 POPULAIRE
5 BON-VIVANT
20 HEUREUX
43 SCRUPULEUX
53 SERIEUX
39 PRUDENT
32 METICULEUX
40 RESERVE
30 MODESTE
42 SCIENTIFIQUE
25 INTELLIGENT
41 RESPONSABLE
35 PERSEVERANT
38 PRATIQUE
19 HABILE
54 SEVERE
44 SEC
15 FROID
36 PESSIMISTE
29 MAUSSADE
24 INSIGNIFIANT
10 ENNUYEUX
11 FIER
58 VANITEUX
34 ORGUEILLEUX
56 TEMERAIRE
2 AUDACIEUX
57 VACILLANT
50 SOUMIS
27 MALADROIT
33 NAIF
9 DERANGE
22 IGNARE
49 SOT
4 BETE
28 MALHONNETE
26 IRRITABLE
17 GROSSIER
6 BRUTAL
3 BESTIAL
16 GASPILLEUR
52 SUPERFICIEL
14 FRIVOLE
31 MEDITATIF
45 SENTIMENTAL
51 SPIRITUEL
18 GENIAL
1 ARTISTE
23 IMAGINATIF
7 CREATIF
12 FIEVREUX
8 DELICAT

Figure 4

Additive Tree obtained from the distance matrix of Table 3. (Qualifiers of the character).

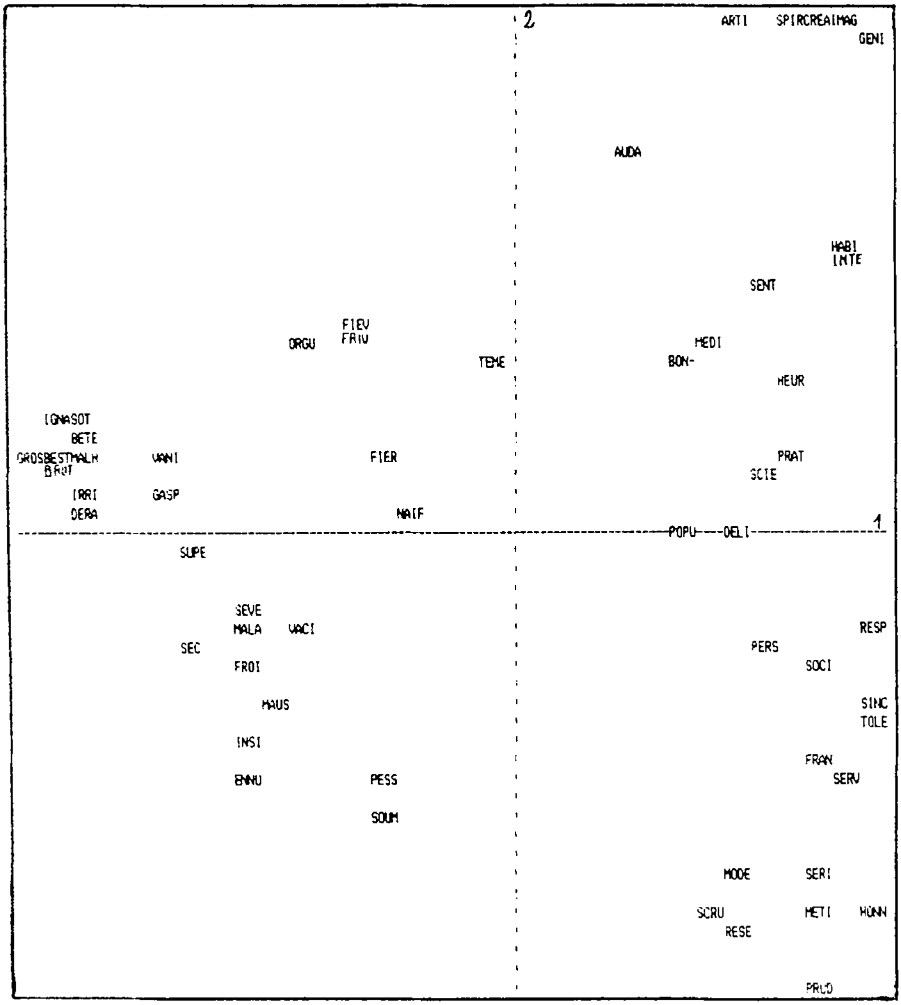


Figure 5

Euclidean representation (factorial analysis of distance) of the distance matrix of Table 3. (qualifiers of the character).
Axe 1 and Axe 2.

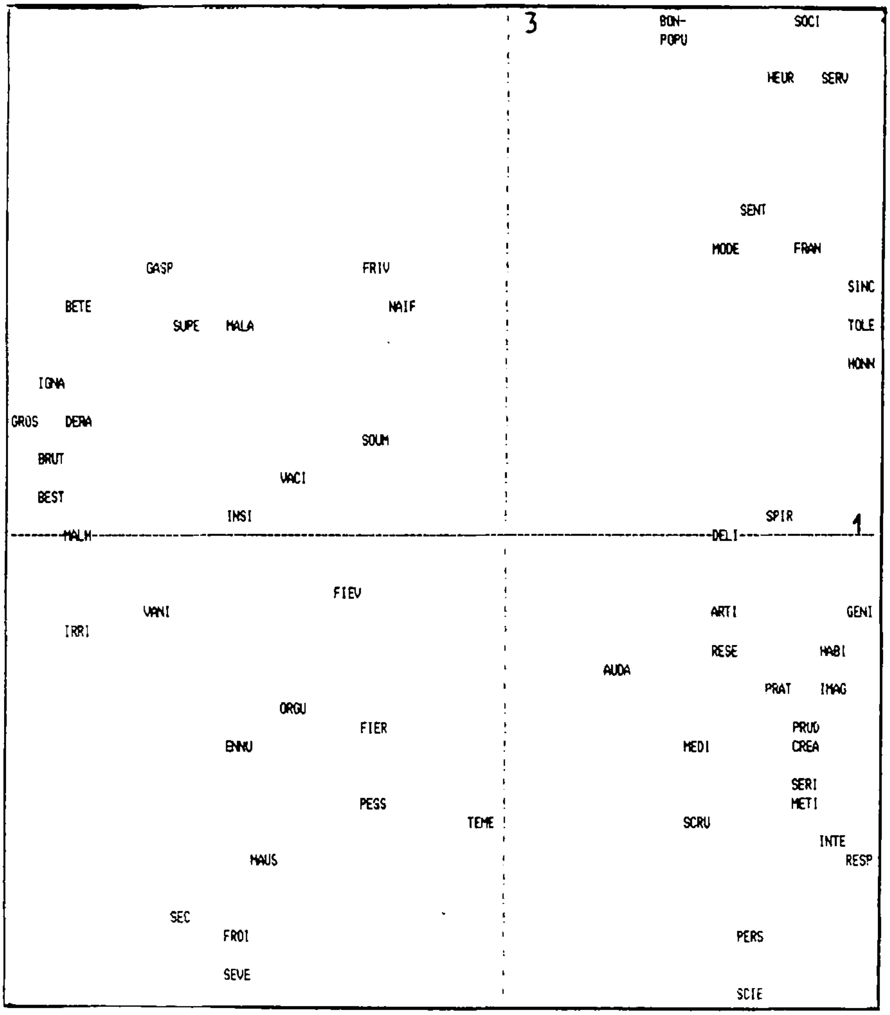


Figure 6

Euclidean representation (factor analysis of distance) of the distance matrix of table 3. (qualifiers of the character). Axe 1 and Axe 3.

	K	M	C	D	U	F	H	R	P	T	N	E	L	O	B	G	V	Q	A	I	J	W	S	Y	I	M	N	O	G	S	H	E	F	R	T	U	W	B	K	A	C	D	L	P	J	Q	V
KNIGHT																																															
MOTHER	36																																														
COOK	39	41																																													
DOCTOR	40	38	47																																												
UMPIRE	38	36	44	44																																											
FISH	1	3	2	2	2																																										
HEDGE	1	2	1	1	1	16																																									
ROOT	1	3	2	2	1	18	40																																								
PLANT	1	2	1	1	1	21	40	42																																							
TREE	1	2	1	1	1	22	41	43	48																																						
NEST	1	2	1	1	.	10	14	12	9	9																																					
EXHAUST	1	.	.	.	.	2	2	1	1	1	5																																				
LABEL	1	1	.	.	.	1	2	1	.	.	14	6																																			
ORNAMENT	1	1	1	.	1	1	3	.	.	.	14	7	18																																		
BLEACH	1	4	3	.	.	1	2	.	.	.	8	12	14	11																																	
GLUE	1	1	1	.	.	1	2	.	.	.	8	12	17	11	40																																
VARNISH	1	.	.	.	.	1	2	.	.	.	8	12	16	14	40	45																															
QUILT	1	2	1	.	.	2	3	2	1	1	17	8	24	24	15	12	13																														
ANCHOR	1	.	.	.	.	4	4	3	.	.	17	9	17	20	13	15	14	23																													
IRON	1	3	2	.	.	1	3	.	.	.	15	8	21	19	24	16	16	27	32																												
JACK	1	.	.	.	.	1	3	.	.	.	15	15	19	19	17	16	16	23	35	41																											
WHEEL	1	.	.	.	.	2	4	1	1	1	13	19	17	16	14	13	13	23	33	34	42																										
SKATE	1	.	.	.	3	3	5	2	1	1	12	8	13	22	8	8	8	18	26	21	20	26																									
YACHT	1	.	.	.	1	5	5	2	1	1	14	7	14	24	8	9	9	19	27	18	18	20	37																								
INCH	.	.	.	.	1	.	.	1	.	.	1	.	9	.	1	1	1	1	1	1	1	1	.	.																							
MEASURE	.	.	.	.	1	.	.	.	.	.	.	.	9	.	.	.	.	.	.	.	.	.	.	.	38																						
NUMBER	.	.	.	.	1	.	.	.	.	.	.	1	10	.	.	.	.	.	.	.	.	.	.	.	37	42																					
ORDER	.	.	.	.	2	.	.	.	.	.	.	.	9	.	.	.	.	.	.	.	.	.	.	.	25	29	30																				
GRADE	1	.	.	.	1	.	.	.	.	.	.	.	11	.	.	.	.	.	.	.	.	.	.	.	27	33	31	44																			
SCALE	.	.	.	.	1	.	.	.	.	.	.	.	9	.	.	.	.	.	.	.	.	.	.	.	29	36	32	42	44																		
HONOR	7	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1	2	3	2																	
EASE	.	.	.	1	1	.	.	.	.	.	.	.	.	2	.	.	.	.	.	1	1	1	4	4	.	1	1	1	1	1	8																
FEAR	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	25															
REGRET	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	25	48														
THRILL	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	10	25	42	42													
URGE	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	20	37	37	35												
WISH	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	10	20	38	37	35	45											
BATTLE	5	.	.	.	1	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1	1	2	2	1	1	1	2	.	.	.	1	.										
KILL	5	.	.	1	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1	1	2	1	1	1	6	2	1	.	.	2	1	40									
AID	2	1	1	2	.	.	.	1	.	.	1	2	1	.	.	1	.	.	.	.	.	.	.	.	.	2	2	3	2	2	2	11	.	.	1	.	.	12	15								
COUNSEL	1	.	.	3	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	2	1	2	2	2	2	9	.	.	1	2	1	5	5	31							
DEAL	.	.	1	2	1	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	1	1	2	1	1	1	1	2	.	.	.	1	2	17	16	13	15						
LABOR	.	.	1	2	1	.	.	.	.	.	.	4	.	.	.	.	.	.	.	1	1	1	.	.	.	2	1	2	2	2	4	7	.	.	1	1	.	17	18	19	11	22					
PLAY	.	.	.	.	4	.	.	.	.	.	.	1	.	4	.	.	.	.	.	.	.	.	7	6	.	2	1	2	2	2	3	7	.	.	3	1	.	16	17	13	6	12	26				
JOKE	.	.	.	1	1	.	.	.	.	.	.	1	.	4	.	.	.	.	.	.	.	.	3	3	1	1	3	2	1	1	1	7	1	.	2	2	3	5	6	10	16	11	5	16			
QUESTION	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	1	1	3	2	1	1	2	3	2	2	2	6	6	5	5	10	18	11	5	6	32		
VOW	1	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1	1	3	2	1	1	8	2	4	3	2	8	10	4	5	9	15	11	3	4	26	32	
YIELD	.	.	.	1	.	1	2	1	1	1	2	10	4	3	1	2	2	2	1	2	2	2	1	2	6	7	8	3	4	5	.	.	.	.	.	.	.	5	6	6	4	15	17	2	3	4	3

Table 4

Co-occurence matrix for 48 English nouns. Data from Miller (1969). The cells indicate the number of subjects who put the words in the same pile (number of subjects: 50).

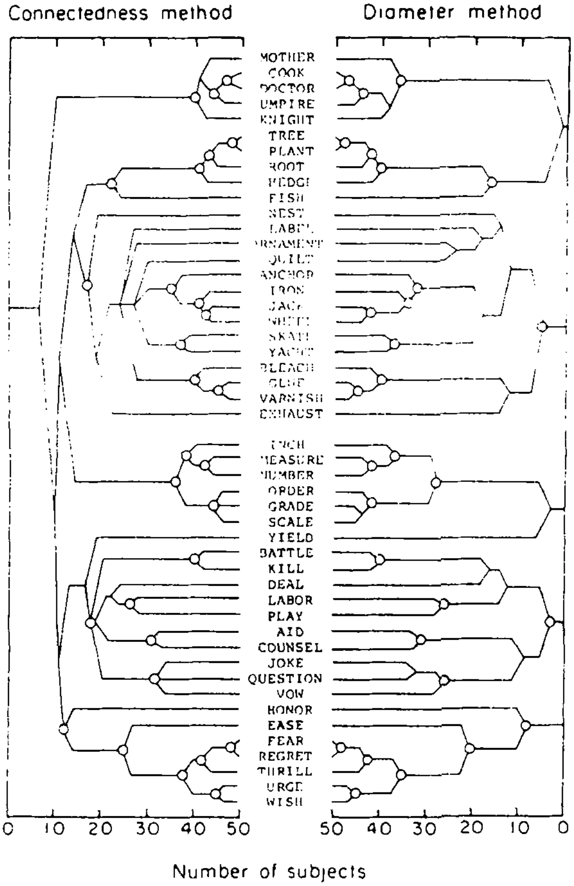

Connectedness method
Diameter method
MOTHER
COOK
DOCTOR
UMPIRE
KNIGHT
TREE
PLANT
ROOT
HEDGE
FISH
NEST
LABEL
ORNAMENT
QUILT
ANCHOR
IRON
JACK
SHEET
SKATE
YACHT
BLEACH
GLUE
VARNISH
EXHAUST
INCH
MEASURE
NUMBER
ORDER
GRADE
SCALE
YIELD
BATTLE
KILL
DEAL
LABOR
PLAY
AID
COUNSEL
JOKE
QUESTION
VOW
HONOR
EASE
FEAR
REGRET
THRILL
URGE
WISH
0
10
20
30
40
50
50
40
30
20
10
0
Number of subjects

Figure 7

Hierarchical solutions for the data of Table 4. From Miller (1969).

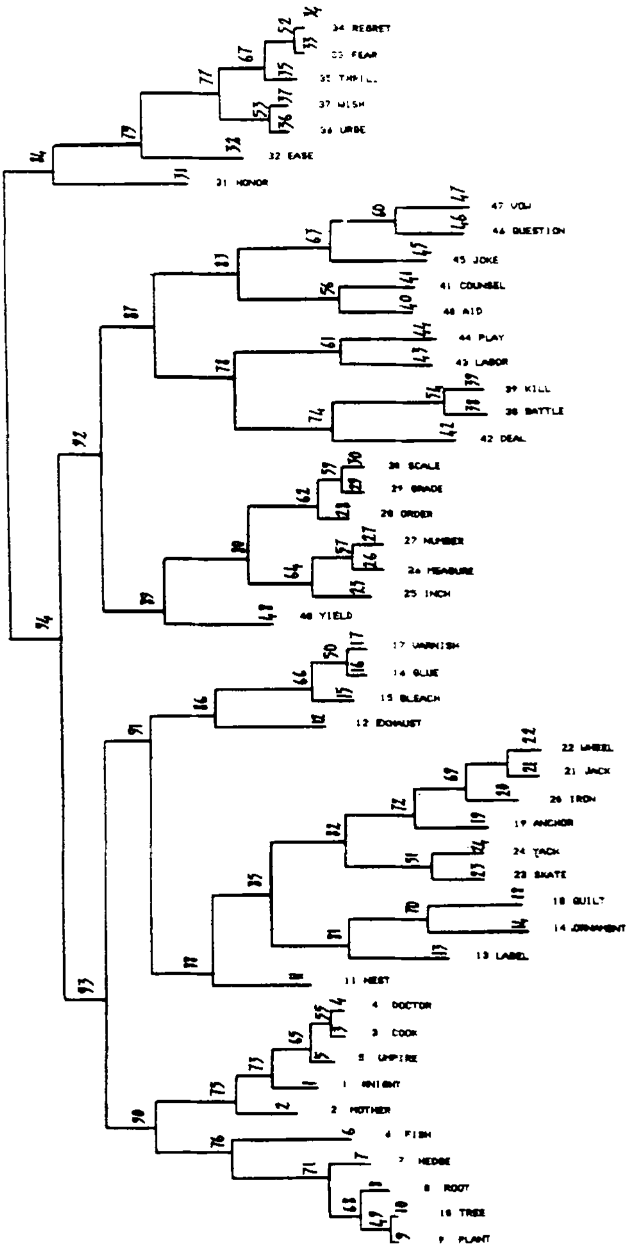

34 REGRET
33 FEAR
35 THRILL
37 WISH
36 URGE
32 EASE
31 HONOR
47 VOW
46 QUESTION
45 JOKE
41 COUNSEL
40 AID
44 PLAY
43 LABOR
39 KILL
38 BATTLE
42 DEAL
30 SCALE
29 GRADE
28 ORDER
27 NUMBER
26 MEASURE
25 INCH
48 YIELD
17 VARNISH
16 GLUE
15 BLEACH
12 EXHAUST
22 WHEEL
21 JACK
20 IRON
19 ANCHOR
24 TACK
23 SKATE
18 QUILT
14 ORNAMENT
13 LABEL
11 NEST
4 DOCTOR
3 COOK
5 UMPIRE
1 INFANT
2 MOTHER
6 FISH
7 HEDGE
8 ROOT
10 TREE
9 PLANT

Figure 8

Additive Tree obtained from the data of Table 4.

	Cat	Cow	Deer	Dog	Goat	Horse	Lion	Mouse	Pig	Rabbit	Sheep
Bear	47.2	27.7	40.1	49.6	19.1	29.0	22.6	29.5	21.4	20.3	16.1
Cat		30.9	56.1	2.0	29.0	25.3	24.1	24.8	43.0	41.5	47.1
Cow			43.6	30.2	11.0	7.7	24.5	34.1	17.0	27.9	8.2
Deer				50.9	44.5	43.0	44.7	39.9	41.1	19.9	53.1
Dog					17.0	24.0	26.9	27.5	45.0	39.4	46.8
Goat						7.2	23.1	39.6	19.5	21.8	1.8
Horse							28.6	32.6	25.7	30.1	15.2
Lion								33.2	29.3	33.3	35.0
Mouse									34.9	22.6	51.9
Pig										25.9	19.6
Rabbit											32.5

Table 5

Proximity matrix between animal terms from Henley (1969).

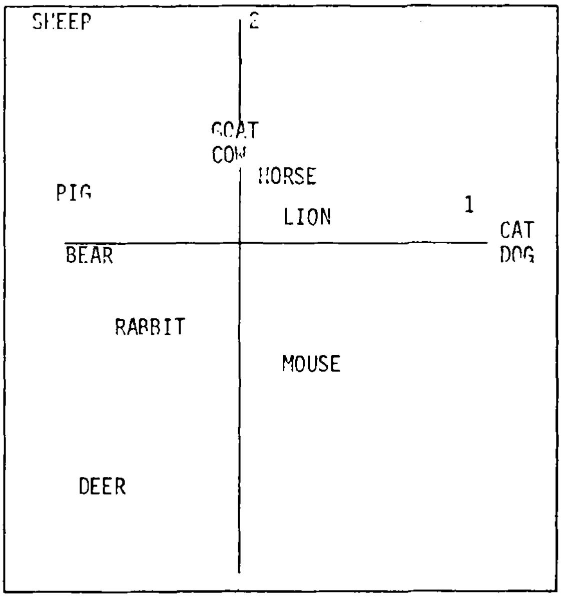


Figure 9

Euclidean representation (factorial analysis of distance) of the distance matrix of Table 5. (Henley's animal terms).

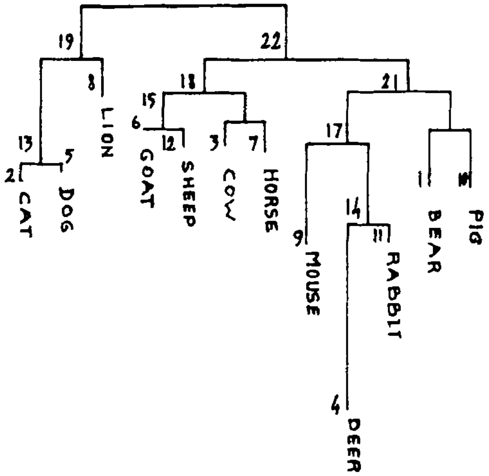

19
22
8
LION
18
21
15
6
12
3
7
17
13
5
2
CAT
DOG
GOAT
SHEEP
COW
HORSE
1
BEAR
PIG
14
9
MOUSE
11
RABBIT
4
DEER

Figure 10

Additive tree obtained from the data of Table 5. (Henley's Animal Terms).

	1[a]	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
dog	×																	
cat	100	×																
lion	91	93	×															
tiger	88	91	100	×														
sheep	86	88	92	91	×													
goat	88	89	94	94	100	×												
face	55	57	57	57	52	51	×											
mouth	62	65	62	61	60	56	95	×										
arm	63	66	65	65	61	60	90	92	×									
finger	62	64	63	63	56	58	94	94	93	×								
leg	58	61	62	62	56	57	89	92	96	90	×							
foot	55	57	59	59	51	53	93	90	92	91	98	×						
celery	69	67	56	56	61	63	60	60	57	62	48	49	×					
spinach	69	68	60	58	65	64	61	62	60	62	56	57	96	×				
lettuce	70	69	60	58	64	64	60	61	59	62	54	55	94	97	×			
potato	68	67	60	59	64	65	65	65	64	67	60	60	87	95	93	×		
rice	66	65	54	53	53	58	66	66	65	70	58	60	88	86	85	92	×	
yam	66	65	57	56	60	62	63	64	62	67	58	57	91	93	92	97	99	×

[a] Numbers indicate item numbers

Table 6

Proximity Matrix between 18 english Nouns obtained by Friendly (1979) in free recall experiments.

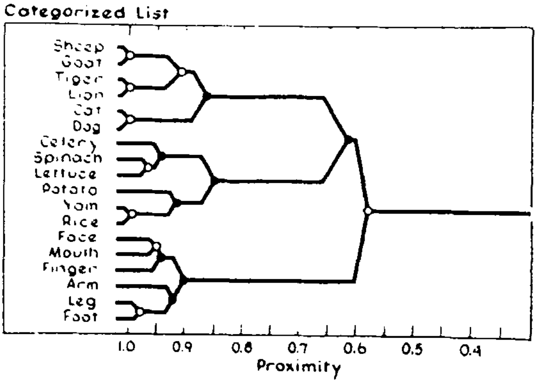

Categorized List
Sheep
Goat
Tiger
Lion
Cat
Dog
Celery
Spinach
Lettuce
Potato
Yam
Rice
Face
Mouth
Finger
Arm
Leg
Foot
1.0
0.9
0.8
0.7
0.6
0.5
0.4
Proximity

Figure 11

Hierarchical solution for the data of Table 6 (from Friendly (1969)).

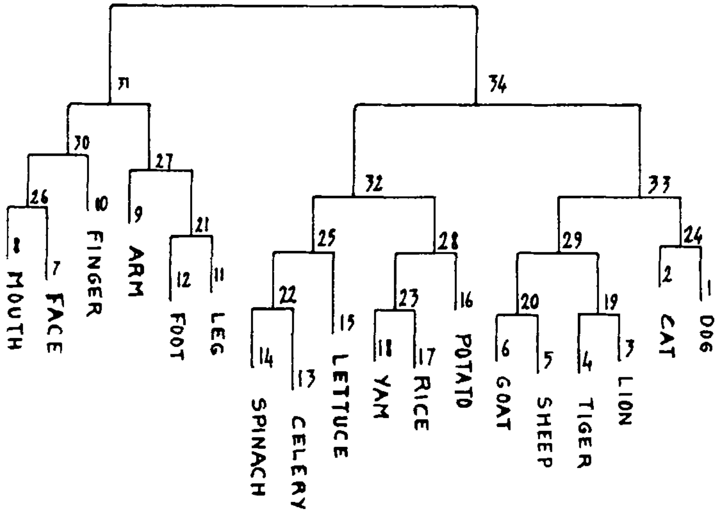

31
34
30
27
32
33
26
10
9
21
25
28
29
24
8
7
12
11
22
15
23
16
20
19
2
1
14
13
18
17
6
5
4
3
MOUTH
FACE
FINGER
ARM
FOOT
LEG
SPINACH
CELERY
LETTUCE
YAM
RICE
POTATO
GOAT
SHEEP
TIGER
LION
CAT
DOG

Figure 12

Additive tree obtained from the data of Table 6 (Friendly's data).

REFERENCES

[1] Abdi, H, Approches préordinales en analyse des données psychologiques, thèse, Aix-en-Provence, (1980).

[2] Abdi, H., Structuration des qualificatifs du caractère, Multigraph, Dijon (1983).

[3] Arabie, P. and Boorman, S., Multidimensional scaling of measures of distances between partitions, Journal of Math. Psychol., 10 (1973), 148-203.

[4] Bartlett, F.C., Remembering, Cambridge University Press, Cambridge (1932).

[5] Benzecri, J.P., L'analyse des données, Dunod, Paris (1973).

[6] Boorman, S.A., Olivier, D.C., Metrics on spaces of finite trees, Journal of Math. Psychol., 10 (1973), 26-59.

[7] Bousfield, W.A., The occurence of clustering in the recall of randomly arranged associates, Journal of General Psychol., 49 (1953), 229-240.

[8] Buneman, P., A note on the metric properties of trees, Journal of Combinatorial theory, 17(b), 48-50.

[9] Carroll, J.D., Chang, J.J., A method for fitting a class of hierarchical tree structure models to dissimilarity data and its application to some body parts data of Miller's, in Proceedings of the 81th Annual Convention American Psychological Association, 8 (1973), 1097-1098.

[10] Colonius, H., and Schulze, R.H., Tree structures for proximity data, Britisch Journal of Math. and Stat. Psychol., 34 (1981), 167-180.

[11] Cunningham, J.P., Finding an optimal tree realization of a proximity matrix, paper presented at the mathematical psychology meeting, Ann Arbor (1974).

[12] Cunningham, J.P., Free trees and bidirectional trees as representations of psychological distances, Journal of Math. Psychol., 17 (1978), 165-188.

[13] Degenne, A., and Verges, P., Introduction à l'analyse de similitude, Revue Française de Sociologie (1973).

[14] Dobson, J., Unrooted tree for numerical taxonomy, Journal of Applied Probability, 11 (1974), 32-42.

[15] Flament, C., L'analyse de similitude, Cahiers du C.R.O., Brussels(1962).

[16] Friendly, M., Methods for finding graphic representations of associative memory structures, in Puff,C.R. (Ed.), Memory organization and structure, Academic Press, New York (1969).

[17] Hakami, S.L., and Yau, S.S., Distance matrix of a Graph and its realizability, Quarterly of Applied Mathematics, 22 (1964), 305-317.

[18] Hartigan, J.A., Representations of similarity matrices by trees, J.A.S.A., 62 (1967), 1140-1158.

[19] Hartigan, J.A., Clustering algorithms, J. Wiley, New York (1975).

[20] Henley, N.M., A psychological study of the semantics of animal terms, Journal of Verb. Learn. and Verb. Behavior, 8 (1969), 176-184.

[21] Hill, M.O., Correspondence analysis, A neglected method, Applied Statistics, 23 (1974), 340-354.

[22] Johnson, S.C., Hierarchical clustering schemes, Psychometrika, 32 (1967), 241-254.

[23] Kerlinger, F.N., Foundations of behavioral research, Holt, New York (1973).

[24] Kruskall, J.B., and Wish, M., Multidimensional scaling, Sage, Beverly Hills (1978).

[25] Luong, X., Voisinage lâche, score et famille scorante, Multigraph, Besançon (1982).

[26] Luong, X., to appear (1983).

[27] Miller, G.A., A psychological method to investigate verbal concepts, Journal of Math. Psychol., 6 (1969), 169-191.

[28] Nishisato, S., Analysis of categorical data: dual scaling and its applications, University of Toronto Press, Toronto (1980).

[29] Pruzansky, S., Tversky, A., and Carroll, J.D., Spatial versus tree representation of proximity data, Psychometrika, 47(1) (1982), 3-24.

[30] Riordan, J., An introduction to combinatorial analysis, J. Wiley, New York (1958).

[31] Rosenberg, S., The method of sorting in multivariate research, in Hirschberg, N., and Humphreys, L.G., (Eds.), Multivariate applications in the social sciences, L. Erlbaum, Hillsdale (1982).

[32] Rosenberg, S., Nelson, C., and Vivekananthan, P.S., A multidimensional approach to the structure of personality impressions, Journal of Pers. and Social Psychol., 9 (1968), 283-294.

[33] Sattath, S., and Tversky, A., Additive similarity trees, Psychometrika, 42(3) (1977), 319-345.

[34] Schiffman, S., Reynolds, M.L., and Young, F.W., Introduction to multidimensional scaling, Academic Press, New York (1982).

[35] Sneath, P.H., Sokal, R.R., Numerical taxonomy, Freeman, San Francisco (1973).

[36] Torgerson, W.S.,Theory and methods of scaling, J. Wiley, New York (1958).

[37] Tulving, E., and Donaldson, W. (Eds.), Organization of memory, Academic Press, New York (1972).

[38] Wegner, D.M., and Vallacher, R.R., Implicit psychology, Oxford University Press, Oxford (1977).

[39] Young, F.W., De Leeuw, J., Takane, Y., to appear: Multidimensional Scaling: theory and methods, L. Erlbaum, Hillsdale.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

TASK-DEPENDENT REPRESENTATION OF CATEGORIES AND MEMORY-GUIDED INFERENCE DURING CLASSIFICATION

Hans Buffart
University of Nijmegen, The Netherlands
Hans-Georg Geissler
Karl-Marx-Universität, Leipzig, Deutsche Demokratische Republik

The formalism outlined preserves the most essential constituents of guided inference.
Processing is specified in terms of operations acting on category representations that include task-constraints and "redefinition of features".
Our rules can be considered as examples of dimensional ordering as discussed by Garner (1978).

1. INTRODUCTION

The problem of how information is processed in classifying objects according to conceptual categories has received much attention throughout the history of experimental psychology. Though this topic ranges over a diversity of experimental paradigms there seems to be some theoretical agreement among researchers that we summarize in the following five assertions.

1.1. REPRESENTATIONS

a) The representation of concepts is based on discrete features.
b) Concepts can be represented in terms of logical functions referring to these features, most commonly by conjunctions and disjunctions of truth values.

1.2. THE CLASSIFICATION-PROCESS

c) There is a distinction between relevant and irrelevant features.
d) The process of classification implies a serial testing of the relevant features or groups of such features.
e) The process of classification comprises an encoding stage translating objects into the feature format for serial processing.

One may devise several paradigms to measure the classification-process partly dependent on the view on this process. For example, a paradigm commonly used in a classification task is to measure reaction-times. It is based on the idea that reaction-time increases with the number of decisions involved. We do not disagree with such a hypothesis in general, but we like to emphasize that the reaction-times do not inform about the character of the decisions and that thus conclusions from experimental results heavily lean upon both the theoretical model and the experimental task.

Many of such models assume that the decision-strategy used by subjects is most parsimonious with respect to the classification to be solved. From such a point of view the strategy can be derived from the logic of the classification problems and reaction-times may be predicted. It is called an 'optimal decision' view. A sàlient example is the serial elimination model proposed by Hick (1952) who considers classification in a multiple choice situation to be based on a sequence of binary decisions. It can be easily proven that with such a binary decision-strategy the number of decisions to be made is minimal.

However, there are two, closely related, problems when using such an assumption of minimal number of decisions on logical grounds independent of a subject's representation of a structure due to 'cognitive constraints' (Note 1). First, such a rule cannot explicitly incorporate task-constraints and thus is not sensitive for changes in task. Instead, several processing models may coexist without any tool to specify in which situation a particular strategy applies. Second, a description of the classification-process of objects purely in terms of a minimal number of decisions is as such not a formulation in terms of the cognitive content of the categories processed, i.e. in order to gather information about the cognitive content from experimental results additional interpretation of the data is needed (cf. Geissler and Puffe (1983)). Pure optimal decision models are without any further assumption not suited to describe task-specificity of human classification performance. A model that accounts for task-specific performance must include the organization of both task-dependent representation and task-dependent processing.

Geissler and Puffe (1983, Note 2) described another view, called the 'guided inference' view. It appeared to account for task-dependence in some experiments that were designed to test crucial differences between this 'guided inference' and the 'optimal decision' view. Despite the limited range of conditions in which the view has so far been subjected to experimental scrutiny we believe that the logics of the approach may be formalized. In the present paper we try to formulate the logics of guided inference in more detail.

2. OPTIMAL DECISION APPLIED

The basic characteristics of guided inference can at best be introduced by discussing its application within the framework of a single experimental paradigm published elsewhere (Geissler and Puffe (1983)) and comparing it with the application of optimal decision. The paradigm and the application of optimal decision will be discussed in this section.

2.1. THE EXPERIMENTAL PARADIGM

In the experiments a hierarchical partition of categories was employed. As an example a three-level partition is depicted in Figure 1 where a partitioning is symbolized by the growing of the number of indices. The paradigm consisted of two tasks. One is commonly known as a verification task and another as a multiple choice reaction, or classification task. In verification a category-name, for instance, "A_2", was given and subjects had to decide whether an object presented belongs to it or not. They then responded with "Yes" or "No" (see also Figure 2). In classification subjects had to respond with the correct category-name of the stimulus presented. The category-names were syllables, known through prior learning, corresponding with the capitals with zero, one or two indices in Figure 1. In classification it was necessary to present first a cue, also known through prior learning, indicating the level, top medium or bottom, on which subjects had to respond (see also Figure 2). A category could consist of several objects, but at the b-level they consisted of one object only. The cases A, A*, B and C in Figure 1 show four types of abstract feature-descriptions of objects used. Figure 2 shows some visual stimuli that fulfil such descriptions: the a_i indicate an outer-contour, the b_i an inner-contour and the c_i a filling-in of the inner-contour.

This experimental paradigm guarantees a strict control of the effects of task- and stimulus-variation at each category-level and allows for a unique and simple description of the data in terms of the features. The data to be discussed here are reaction-times of correct answers. They are assumed to reflect the mean number of decisions to be made in each case. It is

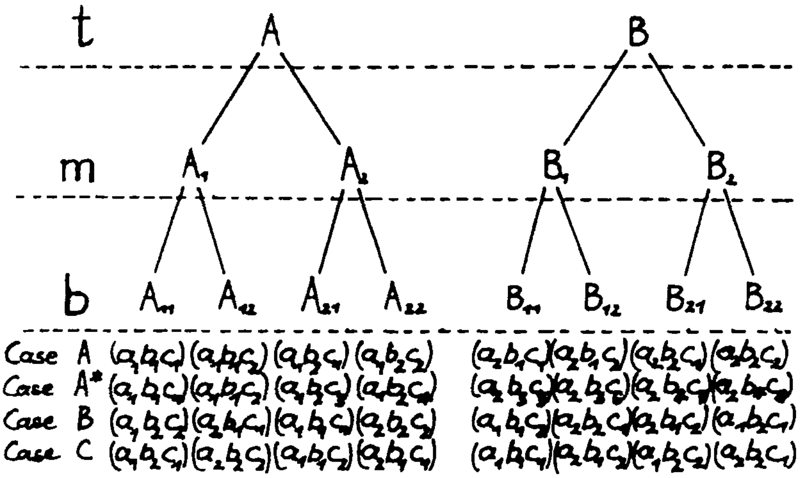

t
A
B
m
A1
A2
B1
B2
b
A11
A12
A21
A22
B11
B12
B21
B22
Case A (a1b1c1)(a1b1c2)(a1b2c1)(a1b2c2) (a2b1c1)(a2b1c2)(a2b2c1)(a2b2c2)
Case A* (a1b1c1)(a1b1c2)(a1b2c3)(a1b2c4) (a2b3c4)(a2b3c3)(a2b4c4)(a2b4c4)
Case B (a1b2c2)(a2b1c1)(a1b1c1)(a2b2c2) (a1b1c2)(a2b2c1)(a2b1c2)(a1b2c1)
Case C (a1b2c1)(a2b2c2)(a1b1c2)(a2b1c1) (a1b2c1)(a2b1c2)(a1b2c2)(a2b2c1)

Figure 1

A hierarchical partition of categories and the four object combinations used in the experiments.

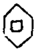
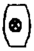
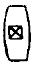
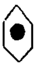
	Cue	Stimulus	Correct response
veri- fication	"A_2"	$(a_1b_2c_1)$ 	"Yes"
	"B_{11}"	$(a_2b_1c_2)$ 	"No"
classi- fication	"t"	$(a_2b_2c_2)$ 	"B"
	"b"	$(a_1b_1c_1)$ 	"A_{11}"

Figure 2

An example of stimuli and answers.

assumed that the parameters to be fitted in order to relate the number of decisions and the reaction-time are independent of the task at hand. Such a requirement of cross-task validation is widely used in perceptual research (Anderson (1976); Birnbaum (1983); Geissler (1980); Geissler, Klix and Scheidereiter (1978); Geissler and Puffe (1983); Geissler, Scheidereiter and Stern (1980)).

2.2. PREDICTIONS BASED ON OPTIMAL DECISION

Conjunctive feature representations

In Cases A and A* each category is related to a set of conjunctively joint features. With the help of Figure 3a the predictions for Case A based on an optimal decision view can be derived in accordance with the assertions mentioned in the introduction. Figure 3a represents a binary decision tree for a serial elimination procedure where each node describes a parallel decision on two features. In each path from top to bottom the letters indicate the relevant conjunctively joint features. If in the classification task the serial testing is assumed to stop at the t-, m- or b-level as indicated by the experimenter, this corresponds with a stop at the 1st, 2nd or 3rd level in Figure 3a when started from the top. The number of feature-tests is the mean number of nodes passed in order to reach the appropriate level, being 1, 2 or 3 respectively. Serial processing in the verification task is equivalent to following a pre-determined top-down path of the tree. The such predicted numbers of decisions for the t-, m- and b-levels are also 1, 2 and 3 respectively. Thus for each hierarchical level of the categories the mean reaction-times for classification and verification are predicted to be equal.

A binary decision in Figure 3a means that a negative decision about a feature, for example a_1, automatically implies a positive decision about the complementary feature at that level, a_2. Therefore, the difference between classification and verification is minimal in the sense that such a binary decision between two features in classification amounts to the same as a Yes-No decision on a single feature in verification. In Figure 3c another decision-tree is depicted that is maximal with respect to the difference between classification and verification. Here the assumption is made that after each negative decision a second test follows checking the complementary feature on a level apart. The mean number of decisions to reach the t-, m- or b-level equals $(1+2)/2 = 1.5$, $(2+3+3+4)/4 = 3$ and

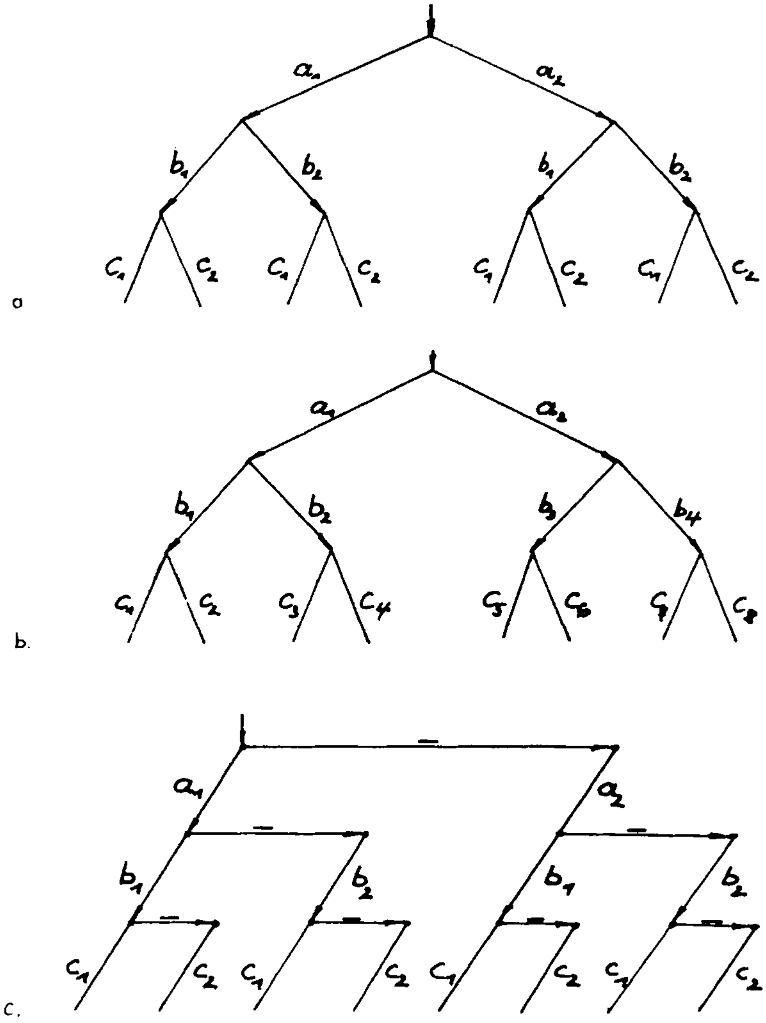

a
a_1
a_2
b_1
b_2
b_1
b_2
c_1
c_2
c_1
c_2
c_1
c_2
c_1
c_2
b.
a_1
a_2
b_1
b_2
b_3
b_4
c_1
c_2
c_3
c_4
c_5
c_6
c_7
c_8
c.
a_1
a_2
b_1
b_2
b_1
b_2
c_1
c_2
c_1
c_2
c_1
c_2
c_1
c_2

Figure 3

Decision-trees for classification, based on the optimal decision hypothesis. a) Case A with minimal number of decisions. b) Case A*. c) Case A with maximal number of decisions.

(3+4+4+5+4+5+5+6)/8 = 4.5 respectively. Thus for each hierarchical level of the categories the ratio of the mean reaction-times for classification and verification is predicted to equal 1.5.

In accordance with the optimal decision hypothesis the reasoning above is based on the conjunctivity not on the feature-outward. This is depicted in the similarity of the decision-trees in Figures 3a and 3b. The latter is a tree for Case A* where at each category-level the relevant features are chosen to be different from any other. The predictions are identical with those for case A*. Figure 4 shows some experimental results for Cases A and A*. The mean reaction-time trends qualitatively contradicts both the minimal and maximal predictions.

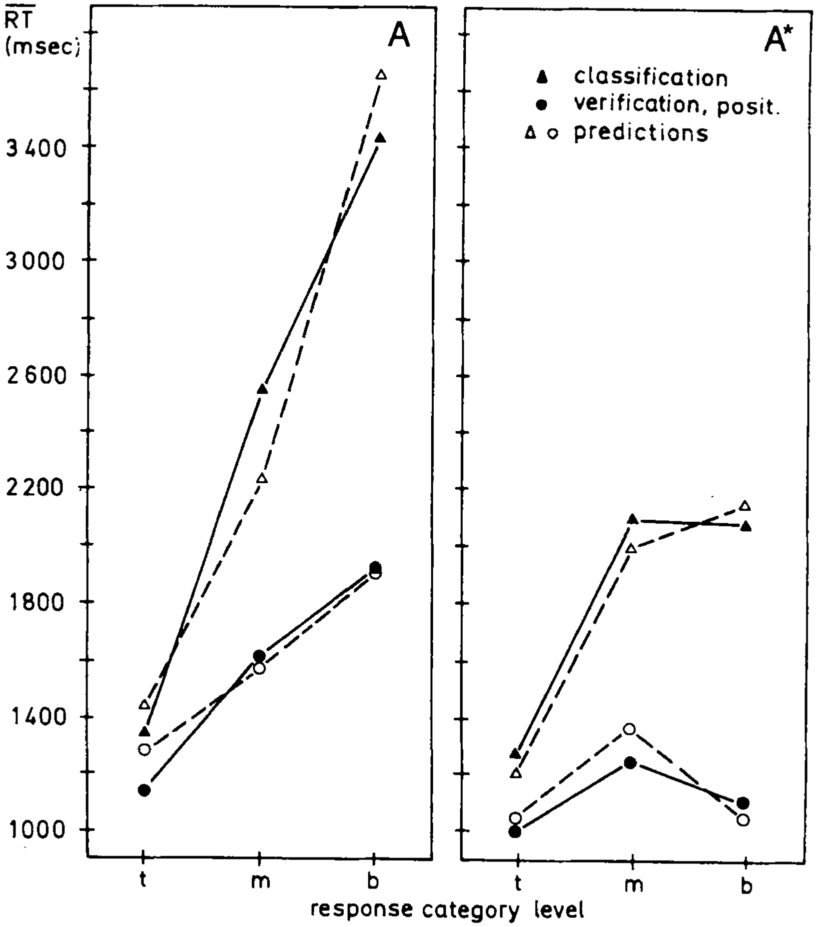


Figure 4

Experimental results (adapted from Geissler and Puffe (1983)) and theoretical predictions.

Disjunctive feature representations

The hierarchical structure in Cases B and C in Figure 1 have been derived from Case A by permuting feature combinations on the b-level, i.e. objects, so that no conjunctive representation on the m- and t-level is possible. From a pure optimal decision view the best strategy for classification and verification is, first, to classify or verify objects by serial elimination on the highest level involving conjunctive representations and, second, to determine, depending on the result on this initial level, the further requested decisions to answer the question.

For Cases B and C the initial level, also called the primary (Hoffmann (1981)) or basic (Rosch (1975)) concept level, is the b-level (cf. Figure 1). Therefore, the mean number of decisions required in the first processing-stage is the mean number of decisions for the b-level in Cases A and A*. For the second processing-stage several implementations are possible. However, responses concerning the m- and the t-level are always obtained through the b-level, so that mean reaction-times at the b-level are never higher than those at the other levels. The minimal classification-verification ratio is obtained by assuming that the decision-tree in Figure 3a is used in the first processing stage and that the second stage scarcely influences reaction-time. At each level the ratio equals 3/3 = 1. The maximal ratio is obtained by assuming in the case of classification that the decision-tree in Figure 3c is used and that each transition from a lower to a higher level in the trees in Figure 1 is a decision-step. At the t-, m- and b-level the ratio equals 6.5/3 = 2.2, 5.5/3 = 1.8 and 4.5/3 = 1.5 respectively.

Figure 5 shows results of an experiment for Cases A, B and C. The results for Case A correspond to those of the former experiment. The results for Cases B and C also disagree with the predictions concerning the differences between verification and classification. In addition, the significant difference between the trends in the Cases B and C contradicts the assumption that classification and verification are independent of feature content. Notice that in Case B the mean reaction-time at the t-level is even much lower than that at the b-level entirely in contrast with the claim that the b-level is a transitional stage for the t-level.

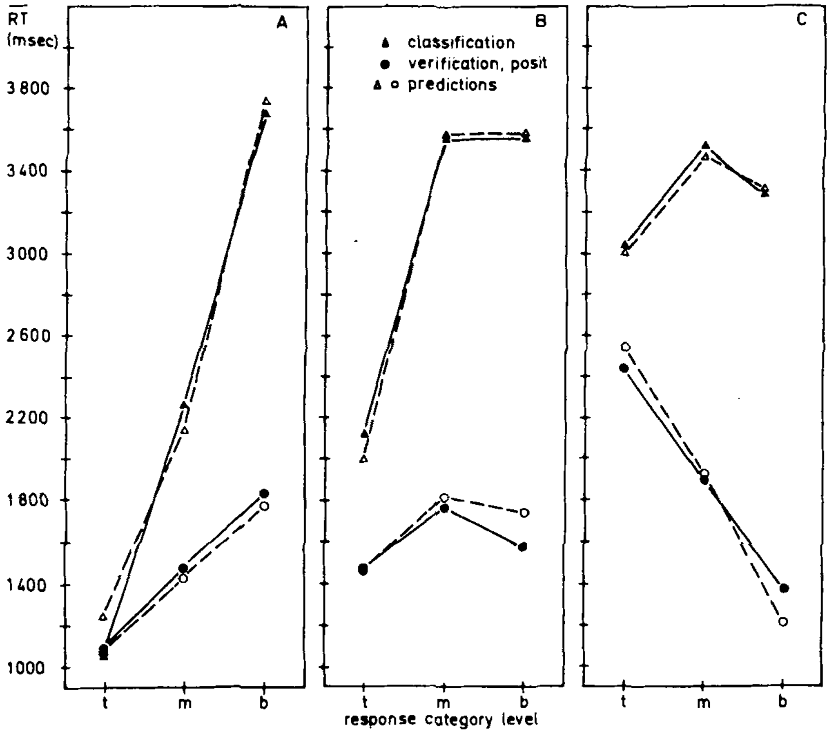


Figure 5

Experimental results (adapted from Geissler and Puffe (1983)) and theoretical predictions.

3. A STRUCTURAL INFORMATION APPROACH TO GUIDED INFERENCE

3.1. GUIDED INFERENCE

In the introduction we already discussed that the optimal decision view cannot account for differences in task and stimulus-content. We believe that the disagreement above is caused by this lack of flexibility. We like to show that classification and verification processes may be interpreted to be determined by the representations of the categories in memory in the *sense that the processes can be completely expressed in terms of representation-parameters*. Such a process is said to be a guided inference process since any particular task can be considered as a problem solving situation where the solution is inferred from the information contained in the representations.

Such a model is suggested by the most salient differences between prediction and results mentioned above. For example:

- The difference observed between verification and classification can be seen as a consequence of task constraints because in verification the category to be checked is already known in advance whereas in classification the correct category is to be found from a set of alternatives. If descriptions of categories would be explicitly involved in processing, verification should expect to be faster than classification.
- In Case A* knowledge of the b- and c-features implies a complete knowledge of the a- and b-features too. In comparison with Case A this could be used to reduce the number of feature tests necessary.
- In Case B for a decision at the t-level it suffices to know only two features, the b_i and c_j, in spite of the disjunctive character of the representations at the m-level. Thus the mean number of decisions at the t-level could be lower instead of higher than this number at the m-level.

3.2. ASSUMPTIONS

A guided inference model requires two types of assumptions. One type concerns the representations and the other type the relation between representations and decisions. We assume that the basic ideas of structural information theory (SIT) about representations (cf. Buffart and Leeuwenberg (1983); Restle (1982)) may also apply here. The assumptions about the decisions are adapted from Geissler and Puffe (1983).

SIT supposes (1) that representations describe identities within and between stimuli and (2) that such descriptions are constrained. The constraints can be expressed in terms of a restricted set of grammar-rules allowing the construction of sometimes intricated expressions. SIT does not make any assumption about constraints on processes, but a recent theoretical study (Buffart, Note 1) suggests the existence of a relation between representation and process constraints. Several representations of a stimulus are allowed. SIT assumes that subjects prefer the representation being most economic including both context and task-demands. This economy- or minimum-principle can be seen as a formal expression of a perceptual stability principle.

A stimulus-representation in the brain is regarded to be less stable as its number of 'free' or non-identical elements increases. Buffart and Leeuwenberg (1983) proposed a memory-model that automatically tends to infer the more stable representation of a stimulus. The minimum-principle has been appeared to be a reasonable tool to calculate the most stable representation.

Geissler and Puffe suppose that object-representations include the category-representations belonging to the task at hand. A representation of an object is decoded into a format that is consistent with its sensory information. Decisions are assumed to be the comparison of the decoded representation with the sensory information. Besides there exist control operations structuring the comparison procedure. They are supposed to originate from the implementation of the categories in the representations. Every operation is assumed to require a fixed time-period, so that a part of the reaction-time may be predicted to be the multiplication of the number of decisions and this period. The other part is supposed to be caused by the encoding process being a constant for each set of stimuli. Because there is no theoretical relation between a representation and its encoding procedure the constant has to be estimated from the data.

3.3. GUIDED INFERENCE APPLIED TO THE DATA

The grammar-rules reflecting the assumptions can at best be introduced at the hand of examples. Their abstract form will be discussed later on. The dashed lines in Figures 4 and 5 show the results of the calculations based on these rules. They coincide with the results of the more intuitively based calculations of Geissler and Puffe (1983). The encoding constants and time-periods for decision are given in Table 1.

time in ms	Figure 4		Figure 5		
	A	A*	A	B	C
decision time	314	314	356	350	306
encoding time	953	731	703	588	694

Table 1

Calculated decision and encoding time in classification and verification.

The basic operator used is a feature connection-operator, \, as introduced in Formulae 1a and 1b. It composes references to features. The first object in Case A, for example, can be described by $a_1 \backslash b_1 \backslash c_1$. The operator is associative.

The operator has the so-called distributive property as defined in SIT (cf. Buffart and Leeuwenberg (1983)). In this paper we use two special forms of this property, called the feature set product rule (Formulae 1c and 1d) and the inner product rule (Formula 1e). The feature set product rule describes generation of objects belonging to a category by full combination of features from sets of mutually exclusive features. The inner product rule describes complete correlation of features. Decoding the left hand-side of a Formula delivers the right hand-side.

$$a\backslash b \Rightarrow ab \tag{1a}$$

$$ab\backslash c \Rightarrow abc \tag{1b}$$

$$(x)\backslash(y_1 \ \dots \ y_n) \Rightarrow x\backslash y_1 \ \dots \ x\backslash y_n \tag{1c}$$

$$(y_1 \ \dots \ y_n)\backslash(x) \Rightarrow (y_1)\backslash(x) \ \dots \ (y_n)\backslash(x) \tag{1d}$$

$$[y_1 \ \dots \ y_n]\backslash[x_1 \ \dots \ x_n] \Rightarrow y_1\backslash x_1 \ \dots \ y_n\backslash x_n \tag{1e}$$

With these rules sets of objects, or of part of objects can be represented. Formula 2 shows a representation of the set of objects in Case A. Formulae 3a-c show it for case A* and Formula 3d represents a set of hierarchical part of these objects. The set of objects in Case B is represented by Formula 4. These representations are the most economic or minimal one. Other representations can easily be constructed.

$$((a_1 \ a_2)\backslash(b_1 \ b_2))\backslash(c_1 \ c_2) \tag{2}$$

$$[(a_1)(a_2)]\backslash[(z_1)(z_2)] \tag{3a}$$

$$z_1 = [(b_1)(b_2)]\backslash[(c_1 \ c_2)(c_3 \ c_4)] \tag{3b}$$

$$z_2 = [(b_3)(b_4)]\backslash[(c_5 \ c_6)(c_7 \ c_8)] \tag{3c}$$

$$[(a_1)(a_2)]\backslash[(b_1 \ b_2) \ (b_3 \ b_4)] \tag{3d}$$

$$([c_2b_2 \ c_1b_1] \, [c_2b_1 \ c_1b_2])\backslash([a_1 \ a_2] \, [a_2 \ a_1]) \tag{4}$$

A pair of brackets {} is used to denote categories in a representation. So $\{x_1y_2\}$ represents a category consisting of one object with features x_1 and y_2 and $\{(x_1x_2)(y_1y_2)\}$ represents a category of four objects described by the product rule inside. If a category consists of two parts connected by the operator \ it suffices to denote the category in one of both parts. For example, $\{x_1y_2\}$ may be denoted by $\{x_1\}\backslash y_2$. The general rule is given in Formulae 5a-f.

$$x\backslash\{y\} \Rightarrow \{x\backslash y\} \tag{5a}$$

$$\{x\}\backslash y \Rightarrow \{x\backslash y\} \tag{5b}$$

$$(\{x\})\backslash y \Rightarrow \{(x)\}\backslash y \tag{5c}$$

$$x\backslash(\{y\}) \Rightarrow x\backslash\{(y)\} \tag{5d}$$

$$[\{x\}]\backslash y \Rightarrow \{[x]\}\backslash y \tag{5e}$$

$$x\backslash[\{y\}] \Rightarrow x\backslash\{[y]\} \tag{5f}$$

The application of these rules and the rules in Formulae 1a-e for the categories "A" and "B" of Case A can be seen in Formulae 6a-f. Formula 6f shows both categories and a one-to-one mapping onto the response-names is possible. Notice that such a one-to-one mapping is also possible from Formula 6a. Therefore, it is not necessary to decode Formula 6a, or in other words for verification or classification it suffices to investigate the "a"-features. However, there is a restriction on such an investigation of features without fully decoding the representation. The investigation can only start at the top or the bottom of the hierarchy of the representation. The hierarchy of the representation in Formula 6a is shown in Figure 3a. Every node is related to a decoding according to one of the Formulae 1c-e if one of the arguments contains a category representation {}. Thus, the investigation of an "a"-feature may start with the investigation of a "c"-feature (bottom) followed by an investigation of a "b"-feature. It also may start with an investigation of an "a"-feature (top), but never with an investigation of a "b"-feature (medium).

$$((\{a_1\}\{a_2\})\backslash(b_1\ b_2))\backslash(c_1\ c_2) \tag{6a}$$

$$((\{a_1\})\backslash(b_1\ b_2)\ (\{a_2\})\backslash(b_1\ b_2))\backslash(c_1\ c_2) \tag{6b}$$

$$(\{(a_1)\}\backslash(b_1\ b_2)\ \{(a_2)\}\backslash(b_1\ b_2))\backslash(c_1\ c_2) \tag{6c}$$

$$(\{(a_1)\backslash(b_1\ b_2)\}\ \{(a_2)\backslash(b_1\ b_2)\})\backslash(c_1\ c_2) \tag{6d}$$

$$\{((a_1)\backslash(b_1\ b_2))\backslash(c_1\ c_2)\}\{((a_2)\backslash(b_1\ b_2))\backslash(c_1\ c_2)\} \tag{6e}$$

$$\{a_1b_1c_1\ a_1b_1c_2\ a_1b_2c_1\ a_1b_2c_2\}\{a_2b_1c_1\ a_2b_1c_2\ a_2b_2c_1\ a_2b_2c_2\} \tag{6f}$$

It is assumed that for a representation given a subject uses such a procedure that a minimal number of decisions is required. This assumption is called the minimal decision hypothesis and it applies after the minimal representation has been chosen. It implies that in Case A subjects start at the top-level and that the reaction-time will be very low. Indeed, this is shown in both Figures 4 and 5. Figure 6a shows how in the case of classification the mean number of decisions is determined. If a subject assumes category "A" he needs one comparison if the feature is a_1 and two comparisons if the feature appears to be a_2. The mean number of decisions equals 1.5. In the case of verification a subject has to investigate whether or not the feature belongs to the category given. For a positive answer precisely one comparison is required.

Formula 7a represents the m-level categories of case A. Again, the comparison-procedure may be started at the bottom- or the top-level of the representation. The latter procedure requires less comparisons. The number of comparisons for classification is shown in Figure 6b. When a subject rightly assumes category "A_1" two comparisons, a_1 and b_1, are necessary. When the category is "A_2" two extra comparisons, a_1 and b_2, are necessary. When the category is "B_1" a subject makes two false comparisons, a_1 and b_1 and two other comparisons, a_2 and b_1. When the category appears to be "B_2" another two comparisons, a_2 and b_2, have to be made. Notice that a comparison-procedure could also have been started with one of the other categories. Once started the procedure follows the hierarchical structure of the representation. Other procedures could have been $A_2A_1B_1B_2$, $A_2A_1B_2B_1$, $A_1A_2B_2B_1$, $B_1B_2A_1A_2$, $B_2B_1A_1A_2$, $B_2B_1A_2A_1$ or $B_1B_2A_2A_1$. The mean number of comparisons equals 4. For a positive answer in a verification-task two comparisons are required.

$$((a_1\ a_2)\backslash(\{b_1\}\{b_2\}))\backslash(c_1\ c_2) \tag{7a}$$

$$((a_1\ a_2)\backslash(b_1\ b_2))\backslash(\{c_1\}\{c_2\}) \tag{7b}$$

Formula 7b represents the b-level categories of Case A. It makes no difference for the number of decisions whether the procedure starts at the top or the bottom of the hierarchy. Figure 6c shows a scheme of the comparison-procedure, that is analogous to the procedure for the m-level. The mean number of comparisons equals 8.5 for classification and 3 for (positive) verification. The calculated mean numbers of comparisons in Cases A, A*, B and C are summarized in Table 2.

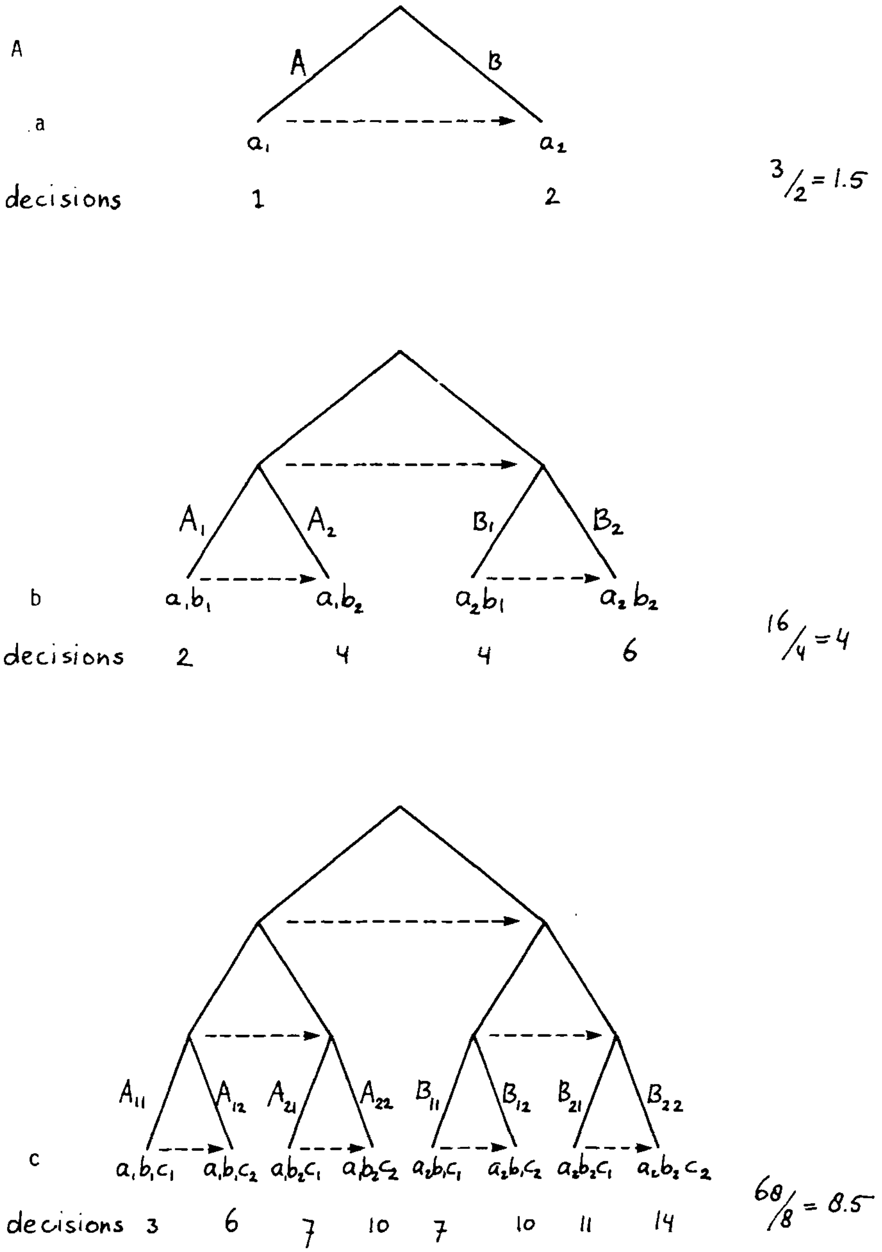

A
a
A
B
a₁
a₂
decisions
1
2
3/2 = 1.5
A₁
A₂
B₁
B₂
b
a₁b₁
a₁b₂
a₂b₁
a₂b₂
decisions
2
4
4
6
16/4 = 4
A₁₁
A₁₂
A₂₁
A₂₂
B₁₁
B₁₂
B₂₁
B₂₂
c
a₁b₁c₁
a₁b₁c₂
a₁b₂c₁
a₁b₂c₂
a₂b₁c₁
a₂b₁c₂
a₂b₂c₁
a₂b₂c₂
decisions
3
6
7
10
7
10
11
14
68/8 = 8.5

Figure 6

Decision-trees for classification in Case A based on the guided inference hypothesis. a) t-level, b) m-level, c) b-level.

mean number of decisions		A	A*	B	C
classification	top	1.5	1.5	4.0	7.5
	medium	4.0	4.0	8.5	9.0
	bottom	8.5	4.5	8.5	8.5
verification	top	1.0	1.0	2.5	6.0
	medium	2.0	2.0	3.5	4.0
	bottom	1.0	1.0	3.0	2.0

Table 2

Calculated number of decisions for classification and verification. Formulae 8ab show the representations of the t- and m-level categories of Case A*. The calculations are identical with the calculations in Case A (see figures 7ab). The representations of the categories at the b-level do not require any discussion. The comparison-procedure is shown in Figure 7c.

$$[(\{a_1\})(\{a_2\})] \backslash [(z_1)(z_2)] \tag{8a}$$

$$[(a_1)(a_2)] \backslash [(\{b_1\}\{b_2\})(\{b_3\}\{b_4\})] \tag{8b}$$

$$([\{c_2b_2\ c_1b_1\}]\,[\{c_2b_1\ c_1b_2\}]) \backslash ([a_1\ a_2]\,[a_2\ a_1]) \tag{9a}$$

$$([c_2b_2\ c_1b_1]\,[c_2b_1\ c_1b_2]) \backslash ([\{a_1\ a_2\}]\,[\{a_2\ a_1\}]) \tag{9b}$$

$$([c_2b_2\ c_1b_1]\,[c_2b_1\ c_1b_2]) \backslash ([\{a_1\}\{a_2\}]\,[\{a_2\}\{a_1\}]) \tag{9c}$$

Formula 9a represents the t-level categories in Case B. The related decision scheme for classification is shown in figure 8a. The calculation of the mean number of decisions follows from the comparison-series:
$b_1c_1 \rightarrow A$, $b_1b_2c_2 \rightarrow A$, $b_1c_1b_2b_1c_2 \rightarrow B$ and $b_1b_2c_2b_1b_2c_1 \rightarrow B$. The mean number of decisions for verification of category "A" follows from the series:
$b_1c_1 \rightarrow$ YES and $b_1b_2c_2 \rightarrow$ YES, implying (2+3)/2 = 2.5 decisions.

Formulae 9bc represent the categories at the m- and b- level. As shown in Figures 8b and 8c the hierarchical structure of both representations is identical so that without any calculation an equal reaction-time is predicted. This agrees with the data (see Figure 5). The comparison-procedure is analogous to the procedure for t-level . The difference between the m- and b-level for verification stems from the fact that a category at the m-level contains two objects. For instance: the two possible comparison-procedures for category A_1, $c_2c_1b_1a_2 \rightarrow$ YES and $c_1b_1a_2c_2 \rightarrow$ YES, imply a mean number of (4+3) = 3.5 decisions.

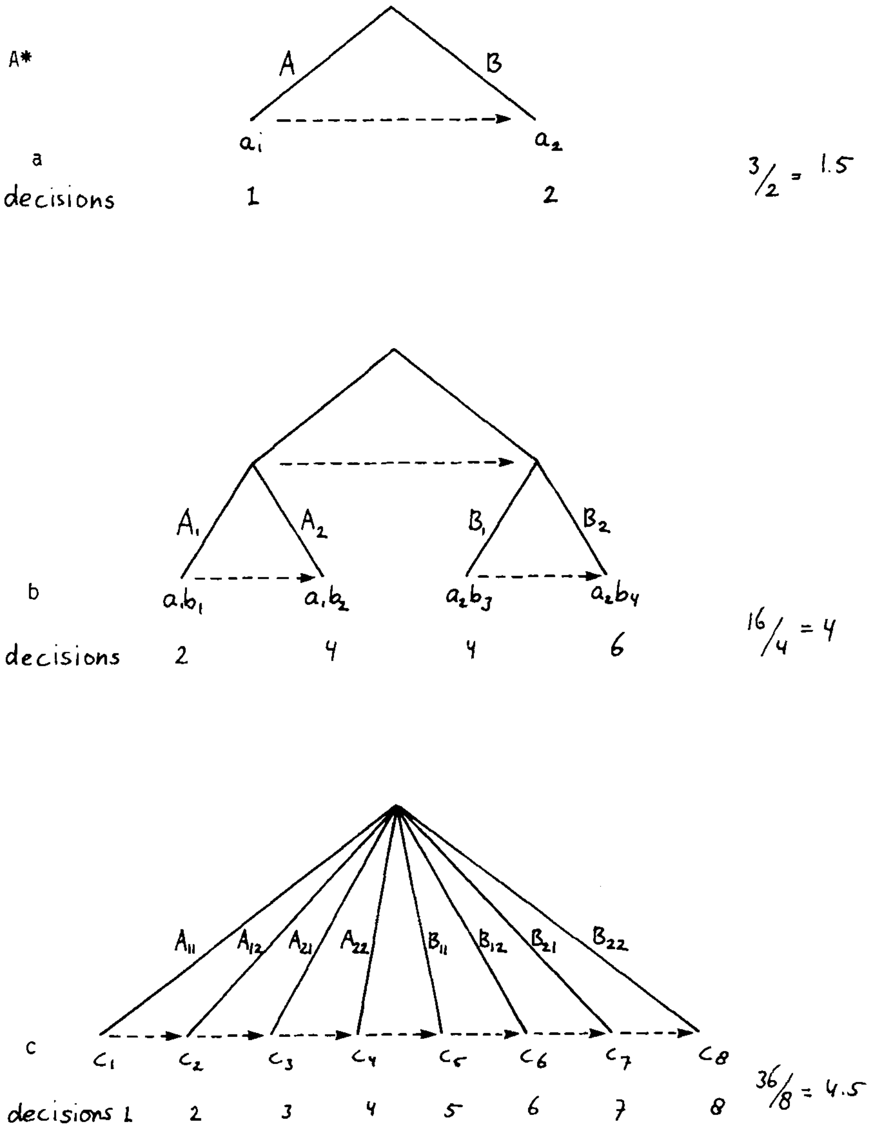

A*
A
B
a
a_1
a_2
decisions
1
2
3/2 = 1.5
A_1
A_2
B_1
B_2
b
a_1b_1
a_1b_2
a_2b_3
a_2b_4
decisions
2
4
4
6
16/4 = 4
A_{11}
A_{12}
A_{21}
A_{22}
B_{11}
B_{12}
B_{21}
B_{22}
c
c_1
c_2
c_3
c_4
c_5
c_6
c_7
c_8
decisions 1
2
3
4
5
6
7
8
36/8 = 4.5

Figure 7

Decision-trees for classification in Case A* based on the guided inference hypothesis. a) t-level, b) m-level, c) b-level.

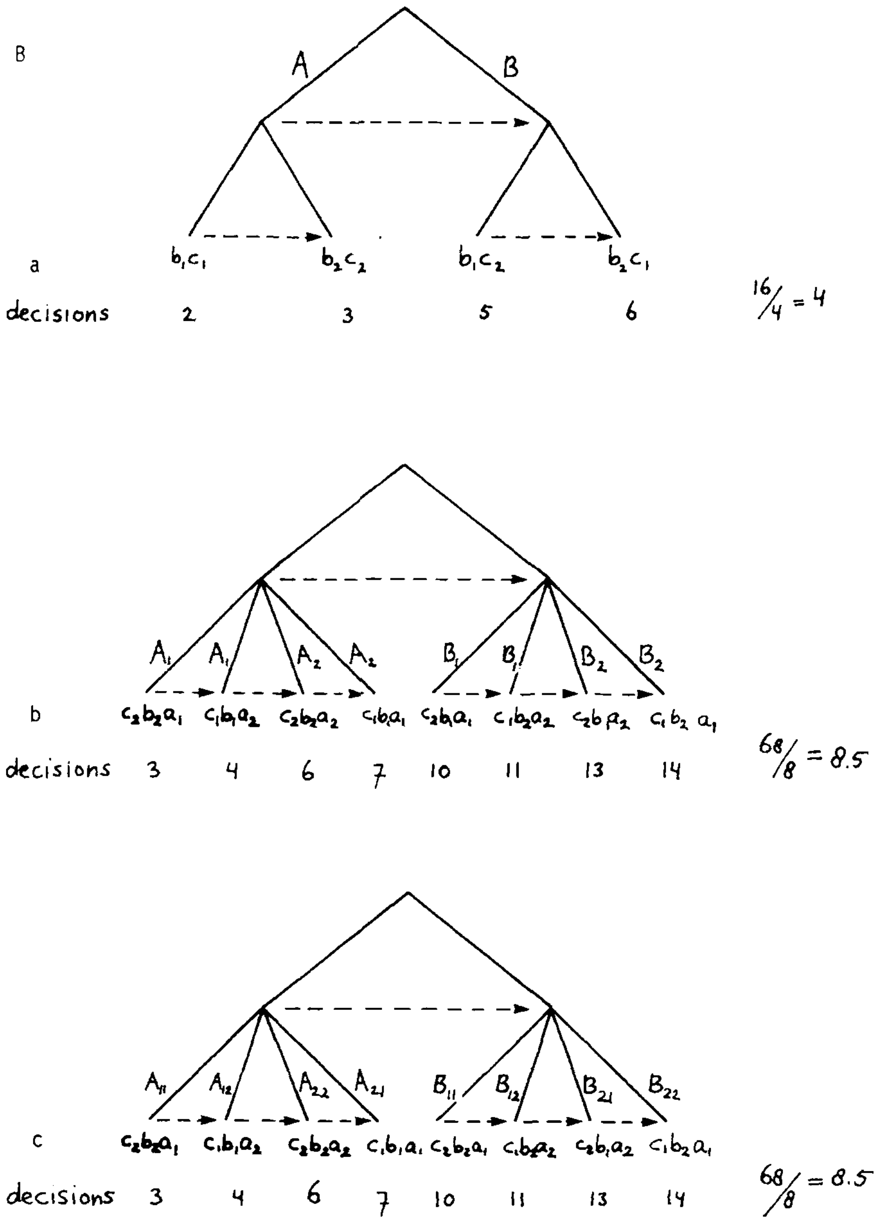

B
A
B
a
b1c1
b2c2
b1c2
b2c1
decisions
2
3
5
6
16/4 = 4
A1
A1
A2
A2
B1
B1
B2
B2
b
c2b2a1
c1b1a2
c2b2a2
c1b1a1
c2b1a1
c1b2a2
c2b1a2
c1b2 a1
decisions
3
4
6
7
10
11
13
14
68/8 = 8.5
A11
A12
A22
A21
B11
B12
B21
B22
c
c2b2a1
c1b1a2
c2b2a2
c1b1a1
c2b2a1
c1b2a2
c2b1a2
c1b2a1
decisions
3
4
6
7
10
11
13
14
68/8 = 8.5

Figure 8

Decision-trees for classification in Case B based on the guided inference hypothesis. a) t-level, b) m-level, c) b-level.

$$(([p_1p_2]\ [p_3p_4]))\backslash(([a_1a_2])\ ([a_2a_1])) \tag{10a}$$

$$([a_1a_2]\ [a_2a_1])\backslash([p_1p_2]\ [p_3p_4]) \tag{10b}$$

$$p_1 \rightarrow b_2c_1,\ p_2 \rightarrow b_2c_2,\ p_3 \rightarrow b_1c_2,\ p_4 \rightarrow b_1c_1 \tag{10c}$$

Formulae 10ab show two representations of the objects used in Case C. Both are *minimal representations. The references p_i refer to a combination of* two features as indicated in Formula 10c. It appeared that a reasonable explanation of the data is only possible if one assumes that subjects regard such a pair of features as one feature. This implies that both features are processed in parallel. Our notation in Formulae 10ab expresses this assumption.

There are also theoretical considerations that support this chunking hypothesis. There is a formal difference between the representations in Case C and the other Cases. The category representations are ambiguous as shown in Formulae 11a-c. One may use the a_i as well as the p_i to represent the categories. Another way of saying is that at the t-level both, non-related, categories "A" and "B" have the same set of features even when the chunking indicated in Formula 10c occurs. We assume that this causes the chunking. It agrees with the basic assumption of SIT (Buffart, Note 1) that elements are chunked when their combination appears to be an element in a representation. It can be proven that in serial representations the distributive rules, such as the rules used in this paper, stem from this assumption.

$$((\{[p_1p_2]\ [p_3p_4]\}))\backslash((\{[a_1a_2]\})\ (\{[a_2a_1]\})) \tag{11a}$$

$$(([\{p_1p_2\}]\ [\{p_3p_4\}]))\backslash(([\{a_1a_2\}])\ ([\{a_2a_1\}])) \tag{11b}$$

$$([\{a_1\}\{a_2\}]\ [\{a_2\}\{a_1\}])\backslash([\{p_1\}\{p_2\}]\ [\{p_3\}\{p_4\}]) \tag{11c}$$

Apart from the chunking there is categorical ambiguity. When during a comparison procedure ambiguity of categorical representations occurs, a subject has to decide which representation is intended. We assume that such a decision is analogous to a feature-comparison so that it takes an equally long time-period. The implementation of this assumption in our formalism is shown in Formulae 12a-e and will be explained later on. Notice that, due to the definitions in formulae 1c-e the indication of a category,{}, in a representation occurs at the right hand-side. At first sight this is not the case in the representations in Formulae 6a, 7a, 8a and 9a, but in these cases the right hand-side is not involved in the comparison-procedure.

Therefore, the brackets at the right hand-side of the Formulae 11a-c are supposed to represent the category intended.

$$\{x\}\backslash\{y\} \Rightarrow \{x\backslash y\} \tag{12a}$$

$$(\{x\})\backslash(\{y\}) \Rightarrow \{(x)\}\backslash\{(y)\} \tag{12b}$$

$$[\{x\}]\backslash[\{y\}] \Rightarrow \{[x]\}\backslash\{[y]\} \tag{12c}$$

$$((\{.....\}))\backslash((\{.....\})) \tag{12d}$$

$$([\{.....\}])\backslash([\{.....\}]) \tag{12e}$$

Formula 12d shows a situation where a subject has to control which the categorical representation is. It is a special case of Formulae 1c-d when the arguments indicate the category. When the arguments are distributed according to Formulae 1c-d one has to keep control which is the category intended. It means that every time when such a hierarchy-node is involved in a comparison-procedure an extra decision has to be taken. This occurs, for example, at the top-level of Formula 11a. With the help of Formula 1c it decodes into Formula 13a. The extra decision, as indicated in Formula 12d, is shown in the hierarchical tree in Figure 9a. The curved arrow indicates that in the case of classification an extra decision is required when a comparison-procedure transfers from category "A" into category "B" or vice versa. Formula 12e indicates a situation analogous to Formula 12d.

$$(\{[p_1p_2]\,[p_3p_4]\})\backslash(\{[a_1a_2]\})\ \ (\{[p_1p_2]\,[p_3p_4]\})\backslash(\{[a_2a_1]\}) \tag{13a}$$

$$\{([p_1p_2]\,[p_3p_4])\}\backslash\{([a_1a_2])\}\ \ \{([p_1p_2]\,[p_3p_4])\}\backslash\{([a_2a_1])\} \tag{13b}$$

$$\{([p_1p_2]\,[p_3p_4])\backslash([a_1a_2])\}\ \ \{([p_1p_2]\,[p_3p_4])\backslash([a_2a_1])\} \tag{13c}$$

Formula 12a and the left hand-side of Formulae 12b-c also indicate a situation where a control-decision has to be taken. In order to create a more transparent formalism such a decision is only attached to Formula 12a, and the left hand-side in Formulae 12b-c is decoded as indicated. Formula 12b is used to decode Formula 13a into Formula 13b. The latter is decoded according to Formula 12a into Formula 13c. Here a control-decision has to be taken as indicated in Figure 9a.

At the hand of Figure 9a the mean number of decisions for classification can be calculated. The possible comparison-series are $p_1a_1 \rightarrow A$, $p_1p_2a_2 \rightarrow A$, $p_1p_2CDp_3a_1 \rightarrow A$, $p_1p_2CDp_3p_4a_2 \rightarrow A$, $p_1a_1p_2CDp_3p_4CDp_1a_2 \rightarrow B$, $p_1p_2a_2CDp_3p_4CDp_1p_2a_1 \rightarrow B$, $p_1p_2CDp_3a_1p_4CDp_1p_2CDp_3a_2 \rightarrow B$, and $p_1p_2CDp_3p_4a_2CDp_1p_2CDp_3p_4a_1 \rightarrow B$, where CD indicates the control-decision.

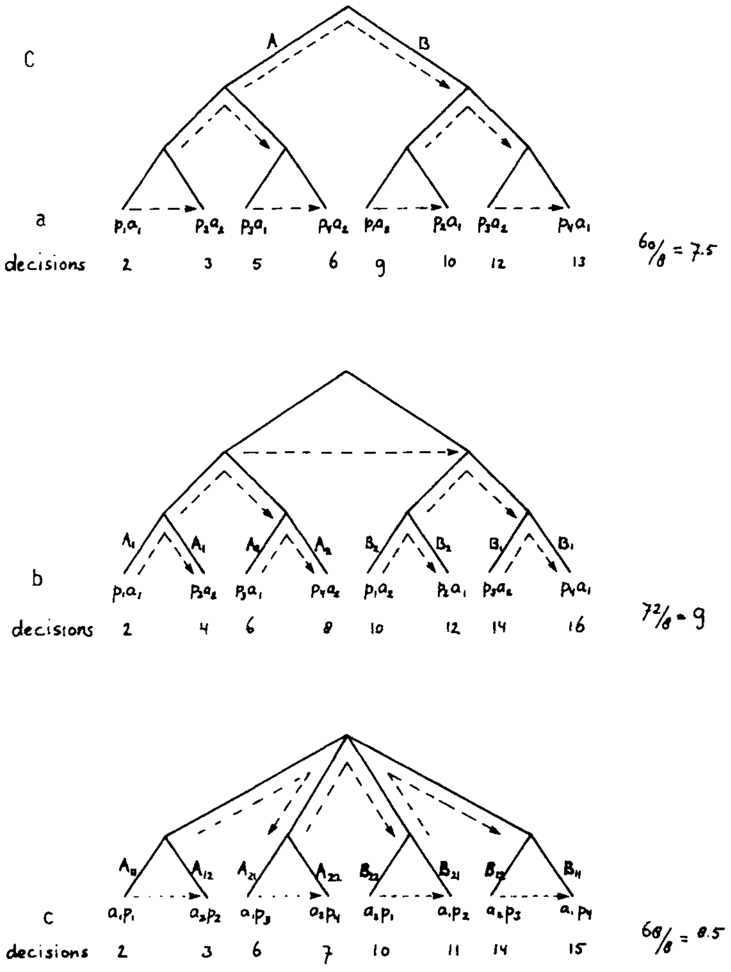

C
A
B
a
p₁a₁ p₂a₂ p₃a₁ p₄a₂ p₁a₂ p₂a₁ p₃a₂ p₄a₁
decisions 2 3 5 6 9 10 12 13
60/8 = 7.5
A₁ A₁ A₂ A₂ B₂ B₂ B₁ B₁
b
p₁a₁ p₃a₂ p₃a₁ p₄a₂ p₁a₂ p₂a₁ p₃a₂ p₄a₁
decisions 2 4 6 8 10 12 14 16
72/8 = 9
A₁₁ A₁₂ A₂₁ A₂₂ B₂₂ B₂₁ B₁₂ B₁₁
c
a₁p₁ a₂p₂ a₁p₃ a₂p₄ a₂p₁ a₁p₂ a₂p₃ a₁p₄
decisions 2 3 6 7 10 11 14 15
68/8 = 8.5

Figure 9

Decision-trees for classification in Case C based on the guided inference hypothesis. a) t-level, b) m-level, c) b-level.

In verification subjects looks for confirmation of features in a fixed set of objects. The object to be tested is arbitrarily chosen, but for every object a control-decision about the category-representation is necessary. The mean number of decisions for verification can be derived from the series $CDp_1a_1 \rightarrow YES$, $CDp_1CDp_2a_2 \rightarrow YES$, $CDp_1CDp_2CDp_3a_1 \rightarrow YES$ and $CDp_1CDp_2CDp_3CDp_4a_2 \rightarrow YES$. The mean number of decisions equals $(3+5+7+9)/4=6$. Because of the ambiguity of the categorical representations in Formulae 11a-c one also could assume that the comparison-procedures are in accordance with the 'reversed' representations. However, then it appears that the mean numbers of decisions in both verification and classification are larger. Our choice for the representations shown in the Formulae is based on the assumption of minimal number of decisions.

4. DISCUSSION

4.1. A SUMMARY OF THE ASSUMPTIONS

Apart from the rules defined, the assumptions that underlie the guided inference approach above can be summarized as follows.

References and features

The basic elements in a representation, called memory-references, refer to features defined in the experiment or to combinations of such features. They represent structures on another level that is not explicitly involved in the task at hand.

Representations

- An object is represented as a combination of references to its features. A set of objects may be represented in terms of rules acting on other rules or references; the decoding of the representation delivers a set of reference combinations uniquely related with the set of objects. A representation expresses the identity of features in a set of objects.
- A category-representation generates the set of objects belonging to the category. It is constructed from a representation of a set of objects by indicating the elements that generate the category during decoding.
- When all category-representations of a set of objects are ambiguous chunking of references occurs if possible.

Decisions

- During classification or verification a subject is supposed to compare references and the related physical shares. The number of comparisons is determined by the subject's representation and the experimental task, but not by the experimentally defined features. The representation is task-dependent.
- The comparison-procedure follows the hierarchical tree generated by the categorical representation. Each decoding of a rule generates a node in the tree. The procedure may start at the top or at the bottom of the tree.
- Each comparison of a reference is supposed to last a standard period of time. It is assumed that references are serially processed and that all structures represented by one reference are processed in parallel. Thus the answer on the question whether serial or parallel processing occurs can be derived from the representations.
- When a categorical representation is ambiguous a subject takes a control-decision every time when the comparison-procedure passes a node in the related hierarchical tree that represents a decoding of a rule with (directly) ambiguous arguments.

The double minimum principle

- The information load of a representation is defined to be the number of references it comprises. From among available alternative representations the ones with the lowest information are supposed to be useful for a subject.
- If more than one minimal representation exists a subject uses from these the representation(s) that requires the minimal number of decisions in classification and verification.

4.2. A POSSIBLE GENERALIZATION

The formalism outlined preserves the most essential constituents of guided inference. Processing is specified in terms of operations acting on category representations that include task-constraints and "redefinition of features". Nevertheless, a further test of the formalism and especially of the representation-rules is required. There is no well-tested set of basic rules available for the description of feature combinations. In the experiments of Hayes-Roth and Hayes-Roth (1977) subjects seem to base their decisions on the power set of feature combinations available. Our rules can be considered as examples of dimensional ordering as discussed by Garner (1978). It cannot be excluded that it will appear that the formulation of a few basic rules is almost impossible because the variability of rules is very large.

This also occurs in SIT. In fact, the theory formulates only a chunking-constraint that roughly implies that, for example, in a sequence ace acf bcf the correlation of a and c as well as the correlation of c and f cannot be explicitly described in one representation. A representation describes either the correlation of a and c or that of c and f. SIT allows all representations that meet this assumption. It supposes that the ultimate choice of the representation is based on a type of stability expressed by the minimum-principle. This choice changes when the context or the task changes (cf. Leeuwenberg and Buffart (1983)). In fact, we also implemented context-influence. Because, we derived the category-representations from the whole set of objects involved in an experiment. If the principles of SIT, formulated for serial structures, are transferable to the domain of conceptual learning the experimental research must give attention to the formulation of the influence of context and task rather than to the formulation of the rules used by the subjects. The paradigm used by Geissler and Puffe (1983) is an example of a measurement of context- and task-influence.

However, as we showed the minimum-principle formulated by SIT is not sufficient to explain the experimental results. We introduced a second minimum-principle concerning the process. It appeared that within the score of the experimental paradigm considered in this paper, it is possible to derive the correct description from a large set of possible theoretical descriptions. From the viewpoint of the double minimum condition inference is guided since first a constraint is put on category representation and only after this a common optimum of representation and processing is selected. It is of high theoretical significance to state the criteria for selecting representations and processing rules in such a way that they remain an invariant under change of paradigm. This is of particular importance since it is directly linked to the difference between the "optimal decision" and "guided inference" approaches. A more general formal definition of the double minimum condition needs a general specification of processing load. Our specification in terms of number of operations and time-periods is certainly not the only one possible.

NOTES

1) The term cognitive constraint has been adapted from Buffart, Leeuwenberg and Restle (1983).

REFERENCES

[1] Anderson, N.A., Integration theory, functional measurement and the psychophysical law, in H.-G. Geissler, and Yu.M. Zabrodin (Eds.), Advances in Psychophysics, Berlin, D.D.R., VEB Deutscher Verlag der Wissenschaften (1976).

[2] Birnbaum, M.H., Scale convergence as a principle for the study of perception, in H.-G. Geissler, H.F.J.M. Buffart, E.L.J. Leeuwenberg, and V. Sarris (Eds.), Modern Issues in Perception, Amsterdam, Berlin, D.D.R., North-Holland Publishing Company, VEB Deutscher Verlag der Wissenschaften (1983).

[3] Buffart, H., and Leeuwenberg, E., Structural information theory, in H.-G. Geissler, H.F.J.M. Buffart, E.L.J. Leeuwenberg, and V. Sarris (Eds.), Modern Issues in Perception, Amsterdam, Berlin, D.D.R., North-Holland Publishing Company, VEB Deutscher Verlag der Wissenschaften (1983).

[4] Buffart, H., Leeuwenberg, E., and Restle, F., Analysis of ambiguity in visual pattern completion, Journal of Experimental Psychology, Human Perception and Performance, 9 (1983), 980-1000.

[5] Garner, W.R., Aspects of a stimulus: features, dimensions, and configurations, in E. Rosch, and B.B. Lloyd (Eds.), Cognition and Categorization, Hillsdale, N.J., Lawrence Erlbaum Associates (1978).

[6] Geissler, H.-G., Perceptual representation of information: dynamic frames of reference in judgment and recognition, in F. Klix, and B. Krause (Eds.), Psychological Research, Humbolt Universität Berlin, 1960-1980, Berlin, D.D.R., VEB Deutscher Verlag der Wissenschaften (1980).

[7] Geissler, H.-G., Klix, F., and Scheidereiter, U., Visual recognition of serial structures, in E.L.J. Leeuwenberg, and H.F.J.M. Buffart (Eds.), Formal Theories of Visual Perception, Chichester, New York, Wiley (1978).

[8] Geissler, H.-G., and Puffe, M., The inferential basis of classification: from perceptual to memory code systems, in H.-G. Geissler, H.F.J.M. Buffart, E.L.J. Leeuwenberg, and V. Sarris (Eds.), Modern Issues in Perception, Amsterdam, Berlin, D.D.R., North-Holland Publishing Company, VEB Deutscher Verlag der Wissenschaften (1983).

[9] Geissler, H.-G., Scheidereiter, U., and Stern, W., Strategies of serial comparison and decision in memory: invariant and task-dependent components, in F. Klix, and J. Hoffmann (Eds.), Cognition and Memory, Berlin, D.D.R., VEB Deutscher Verlag der Wissenschaften (1980).

[10] Hayes-Roth, B., and Hayes-Roth, F., Concept learning and the recognition and classification of exemplars, Journal of Verbal Learning and Verbal Behaviour, 16 (1977), 321-328.

[11] Hick, W.E., On the rate of gain of information, Quarterly Journal of Experimental Psychology, 4 (1952), 11-26.

[12] Hoffmann, J., Concepts, their formation, representation and identification (in German), in Proceedings of the XXnd International Congress of Psychology, Leipzig, D.D.R. (1981), 97-105.

[13] Leeuwenberg, E., and Buffart, H., An outline of coding theory: a summary of related experiments, in H.-G. Geissler, H.F.J.M. Buffart, E.L.J. Leeuwenberg, and V. Sarris (Eds.), Modern Issues in Perception, Amsterdam, Berlin, D.D.R., North-Holland Publishing Company, VEB Deutscher Verlag der Wissenschaften (1983).

[14] Restle, F., Coding theory as an integration of gestalt psychology and information processing theory, in J. Beck (Ed.), Organization and Representation in Perception, Hillsdale, N.J., Lawrence Erlbaum Associates (1982).

[15] Rosch, E., Cognitive representations of semantic categories, Journal of Experimental Psychology, 3 (1975), 382-407.

REFERENCE NOTES

[1] Buffart, H., Seven minus two and structural operations, International Report, University of Nijmegen, The Netherlands (1984).

[2] Puffe, M., Visuelle Erkennung als anforderungsabhängige Inferenz über sequentiellen Kategorierepresentationen, Unpublished Thesis, Berlin, D.D.R., Humbolt Universität (1983).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

OUTLINE OF A THEORY OF BRIGHTNESS, COLOR, AND FORM PERCEPTION

Stephen Grossberg
Center for Adaptive Systems
Boston University

This paper describes new concepts and mechanisms from a real-time visual processing theory that has been used to explain paradoxical data about brightness and form perception. These data include the Craik-O'Brien effect, the Land brightness and color demonstrations, the fading of stabilized images, neon color spreading, complementary color induction, completion of illusory contours, and binocular rivalry. Two functionally distinct contour processes interact to generate these brightness and form properties in the theory. A Boundary Contour process is sensitive to the amount of contrast but not to the direction of contrast in scenic edges. It includes a binocular matching stage that is sensitive to spatial scale, orientation, and binocular disparity, and whose outcome triggers a process of monocular contour completion. These completed contours form the boundaries of monocular perceptual domains. A Feature Contour process is sensitive to both the amount of contrast and to the direction of contrast in scenic edges. It triggers a diffusive filling-in reaction of featural quality within perceptual domains whose boundaries are dynamically defined by the completed boundary contours. The diffusive filling-in reactions take place within syncytia of cell compartments. These preprocessed monocular representations give rise to a percept via a process of binocular resonance. The percept takes the form of standing waves of patterned activity among multiple spatial scales. The Boundary Contour process is hypothesized to be analogous to interactions between the hypercolumns in area 17 of the visual cortex. The Feature Contour process is hypothesized to be analogous to interactions between the cytochrome oxydase staining blobs in area 17 and prestriate cortex in area 18.

1. INTRODUCTION : PARADOXICAL VISUAL DATA AS MANIFESTATIONS OF ADAPTIVE VISUAL MECHANISMS

This chapter outlines some of the new concepts and mechanisms that my colleagues and I have been using to explain paradoxical data about brightness and form perception (Cohen and Grossberg, 1983a, 1983b, 1984; Grossberg, 1981, 1983a, 1983b; Grossberg and Mingolla, 1984). That paradoxical data abound in the field of visual perception does not seem as surprising when one considers the manner in which visual information is acquired. Light passes through retinal veins before it reaches retinal photoreceptors, and light does not influence the retinal region corresponding to the blind spot. The percepts of human observers are not distorted, however, by their retinal veins or blind spots during normal viewing conditions. Thus some images that are retinally present are not perceived because our visual processes are adaptively designed to free our percepts from imperfections of visual scanning. These adaptive visual processes can also generate paradoxical percepts, as during the perception of stabilized images (Riggs, Ratliff, Cornsweet, and Cornsweet , 1953 ; Pritchard, Heron and Hebb, 1960 ; Pritchard , 1961 ; Yarbus , 1967), and of filling-in reactions (Gerrits and de Haan, 1966; Gerrits and Timmerman, 1969; Gerrits and Vendrick, 1970).

Other adaptive visual mechanisms exist than those which compensate for retinal veins and the blind spot. The visual mechanisms that prevent perception of retinal veins and the blind spot are operative even when only one eye is open. Due to the action of binocular visual mechanisms, some retinal images that can be monocularly perceived may not be perceived during binocular viewing conditions. Binocular rivalry provides a classical example of this fact (Blake and Fox, 1964; Cogan, 1982; Kaufman, 1974; Kulikowski, 1978). The properties of binocular rivalry also seem less paradoxical when one considers how the suppression of some monocular data can achieve a more consistent depthful binocular representation of a scene (Kaufman, 1974; Grossberg, 1980, 1983a).

The above examples illustrate that retinally present images may be suppressed to achieve more consistent visual representations of the external world. For a similar reason, observers may see images that are not retinally present. The remarkable Land color demonstrations (Land, 1977) are, for example, often interpreted as consequences of the visual system's ability to process the reflectances of perceived objects despite the variability of the lighting conditions under which the objects are observed. Our theory clarifies the adaptive mechanisms that give rise to the fading of stabilized images, filling-in, and the Land brightness color demonstrations. The same mechanisms also shed light on the perception of illusory figures (Gellatly, 1980; Kanizsa, 1974; Kennedy, 1978, 1979, 1981; Parks, 1980; Parks and Marks, 1983; Petry, Harbeck, Conway, and Levey, 1983), of neon color spreading (Redies and Spillman, 1981; van Tuijl, 1975; van Tuijl and de Weert, 1979; van Tuijl and Leeuwenberg, 1979), and of Julesz stereograms (Julesz, 1971; Kaufman, 1974).

Every theory of adaptive behavior is faced with the problem of specifying precisely the sense in which its targeted behavior is adaptive, without becoming circular or falling into an infinite regress. This difficult problem does not go away just by citing platitudes about Darwinian evolution. The principles of organization and real-time mechanisms that instantiate our view of adaptive behavioral processes are often far removed from teleological concepts that are easily expressed using concensual language. Some of our concepts analyse, for example, how the nervous system, or indeed any cellular machine, may respond to its input environment in a sensitive and stable fashion without losing its capability for regulated self-organization (Grossberg, 1978, 1980, 1982). This approach indicates how behavioral properties which seem to reflect an intelligent teleology may arise as the epiphenomena of general constraints on the possibility of doing accurate information processing at all.

The present chapter does not derive these principles and mechanisms from first principles. Rather it outlines how some of these general tools, once available, can be specialized for use in brightness and form perception. My focus will be on processes that are hypothesized to occur at a central level of neural processing, notably in the visual cortical areas 17 and 18. The neural interpretation of these processes is not, however, critical to understanding how they work. All of the processes may be interpreted as formal rules for real-time visual information processing. The paper with Professor Carpenter (Carpenter and Grossberg, (1981), by contrast, illustrates how an adaptive approach may be used to clarify data taken from the most peripheral photoreceptive cells of the vertebrate nervous system. This work may also be studied as a purely formal model, notably a model concerning how a receptive device can recalibrate its sensitivity in response to an input source whose intensity can fluctuate over a very large range through time.

2. BOUNDARY CONTOURS AND FEATURE CONTOURS.

One of the most important new concepts of our theory is that several types of contour extracting processes are simultaneously operative during brightness and form perception (Grossberg, 1983a, 1983b). Two of these contour processes are said to generate boundary contours and feature contours, respectively. The very fact that more than one type of contour process is hypothesized to exist implies that care must be exerted in analysing what one means by a contour, or edge, and how it is computed. For example, neither of our contour processes is computed using zero crossings (Marr and Hildreth, 1980), which are wholly inadequate for our purposes (Grossberg, 1983a). This chapter will describe and use these new contour concepts in an intuitive and nontechnical way.

Before illustrating how boundary and feature contour processes help to explain paradoxical visual data, I will differentiate them using several of their more salient properties.

2.1. BOUNDARY CONTOURS AND CONTOUR COMPLETION

(i) Contrast : A boundary contour is sensitive to the amount of contrast, but not to the direction of contrast, at an edge of a visual scene. Thus a boundary contour can be activated by either a light-dark edge or a dark-light edge at a fixed scenic position.

(ii) Completion : Several boundary contour activations that correspond to contrast differences at different scenic positions may interact via a process of contour completion. The contour completion process is capable of synthesizing global visual contours from local contour fragments. The contours that are perceived in illusory figures are assumed to be generated by this contour completion process. Completed boundary contours are used to bound perceptually significant domains in the theory; hence the name boundary contour.

A simple demonstration of a contour process with these properties can be made as follows (Grossberg and Mingolla, 1984). Divide a square into two equal rectangles along an imaginary boundary. Color one rectangle a uniform shade of gray. Color the other rectangle in shades of gray that process from light to dark as one moves from end 1 of the rectangle to end 2 of the rectangle. Color end 1 a lighter shade than the uniform gray of the other rectangle, and color end 2 a darker shade than the uniform gray of the other rectangle. Then as one moves from end 1 to end 2, an intermediate gray region is passed whose luminance approximately equals that of the uniform rectangle. At end 1, a light-dark edge exists from the nonuniform rectangle to the uniform rectangle. At end 2, a dark-light edge exists from the nonuniform rectangle to the uniform rectangle. Despite this reversal in the direction of contrast from end 1 to end 2, an observer can see an illusory edge that joins the two edges of opposite contrast and separates the intermediate rectangle regions of equal luminance.

This demonstration illustrates the existence of a process of completing contours between edges with opposite directions of contrast. This contour completion process is thus sensitive to amount of contrast, but not to direction of contrast.

(iii) Binocular Matching : A monocular boundary contour can be generated when a single eye views a visual scene. When two eyes view a visual scene, pairs of monocular boundary contours become engaged in a process of binocular matching. This binocular matching process is sensitive to such factors as spatial scale, binocular disparity, and edge orientation. Due to the binocular matching process, a fused binocular boundary contour may be generated from a pair of matched monocular boundary contours. Alternatively, one of the monocular boundary contours may be suppressed, or a double image of unfused monocular boundary contours may occasionally be generated. These outcomes provide the local substrate for such global processes as binocular rivalry and depth perception in the theory (Grossberg, 1980, 1983a). In particular, if monocular boundary contours corresponding to the left eye are momentarily suppressed by binocular mismatch, then these monocular boundary contours cannot contribute to the contour completion process corresponding to the left visual field. Visual percepts that can be seen during monocular viewing may thus be suppressed by this binocular matching process. These processes of binocular matching and contour completion are interpreted neurally to occur in the striate cortex (area 17).

The boundary contour processes are like frames without pictures. The pictorial data themselves are derived from the feature contour processes. This suggestive analogy with pictures and frames should not be taken too literally, however, since the same visual sources input to both the boundary contour process and the feature contour process, and the outputs of both types of processes interact in a contex-sensitive way to generate visual representations.

2.2. FEATURE CONTOURS AND DIFFUSIVE FILLING-IN

(i) Contrast : The feature contour process is sensitive to direction of contrast as well as to amount of contrast, unlike the boundary contour process. For example, it is obviously important in computing the relative brightness across a scenic boundary to take account of which side of the scenic boundary has a larger reflectance. A similar remark holds at a red-green scenic boundary.

(ii) Completion : Boundary contours activate a contour completion process that generates the global contours which bound monocular perceptual domains. Feature contours activate a diffusive filling-in process that spreads featural qualities, such as brightness or color, across these monocular perceptual domains. Such a diffusive filling-in reaction is, for example, hypothesized to prevent perception of the blind spot.

This diffusive filling-in process is spatially limited by boundary contours. The binocular matching process that occurs between monocular boundary contours chooses among the possible spatial scales that will receive boundary contour signals. Thus, although a feature contour may be activated across all of these spatial scales, only those scales that receive boundary contour signals will generate barriers to the diffusion of featural quality. Only these scales will be capable of contributing to the final percept in the theory, as the subsequent discussion will show.

(iii) Monocular Filling-In : Because monocular boundary contours are binocularly matched before giving rise to completed contours, each visual field (left and right) may process different completed boundary contours, just as in binocular rivalry. These distinct completed boundary contours can differentially regulate the monocular processing of featural quality in their respective visual fields.

These concepts will now be refined using a macrocircuit diagram and applied to several types of perceptual phenomena.

3. MACROCIRCUIT OF PROCESSING STAGES.

Figure 1 describes a macrocircuit of processing stages that is capable of synthesizing global depth, brightness, and form information from monocularly

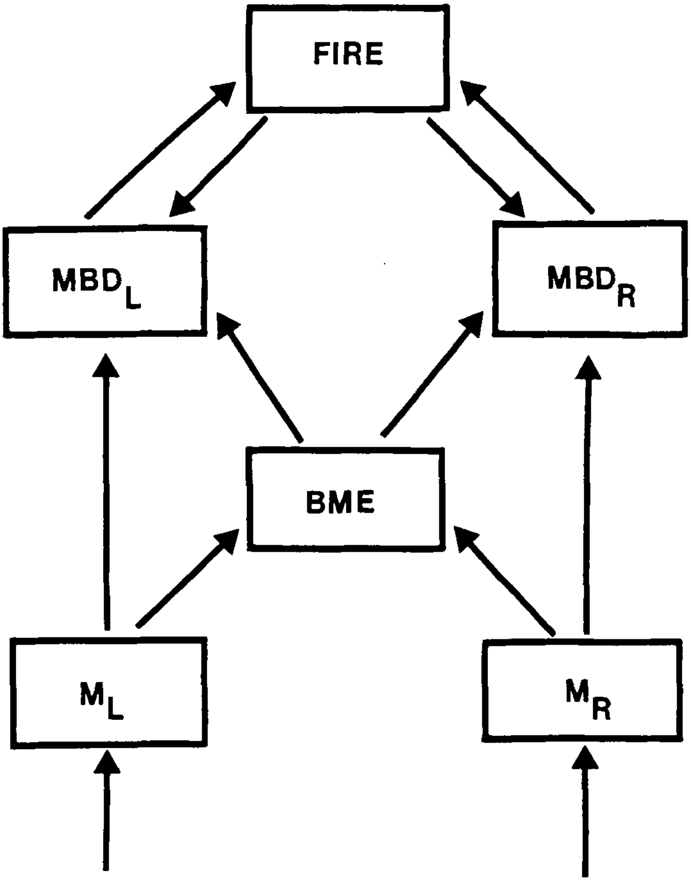


Figure 1

Network stages used to extract global depth, brightness, and form information from monocular and binocular images. Table I lists the names of these stages and their possible neural analogs.

and binocularly viewed patterns. Table I lists the names of the abbreviated processing stages, as well as the neural structures that presently seem most likely to execute analogous processes.

Table I

M_L - Left Monocular Representation (Lateral geniculate nucleus)

M_R - Right Monocular Representation (Lateral geniculate nucleus)

BME - Binocular Matching of Edges (Hypercolumn interactions in striate cortex - Area 17 (Hubel and Wiesel, 1977))

MBD_L - Left Monocular Brightness Domain (Cytochrome oxydase staining blobs - Area 17 - and/or prestriate cortex - Area 18 (Horton and Hubel, 1981; Hendrickson, Hunt, and Wu, 1981; Hubel and Livingstone, 1981; Livingstone and Hubel, 1982))

MBD_R - Right Monocular Brightness Domain (Prestriate cortex - Area 18 - and/or cytochrome oxydase staining blobs in Area 17)

FIRE - Filling-In Resonant Exchange (Prestriate cortex - Area 18)

The BME stage generates oriented monocular boundary contours in response to inputs from M_L and M_R, binocularly matches these monocular boundary contours, and monocularly completes the boundary contours that survive the binocular matching process. The completed left boundary contours generate scale-specific boundary signals to the MBD_L stage. The completed right boundary contours generate scale-specific boundary signals to the MBD_R stage.

Monocular feature contours are generated by $M_L \rightarrow MBD_L$ signals and $M_R \rightarrow MBD_R$ signals. Thus the monocular regions M_L and M_R give rise to both boundary contour signals and to feature contour signals. The boundary contours undergo binocular matching and monocular contour completion before rejoining the corresponding feature contours at MBD_L and MBD_R. This divergence of feature contour processing from boundary contour processing allows the boundary contours to benefit from scale-specific, orientation-specific, and disparity-specific binocular selection within BME before the surviving boundary contours are passed among the spatial scales of MBD_L and MBD_R.

4. DIFFUSIVE FILLING-IN WITHIN MONOCULAR BRIGHTNESS AND COLOR DOMAINS.

The monocular boundary contour signals from stage BME to stage MBD_L define barriers to the lateral diffusion of featural activity that is triggered by feature boundary signals from M_L to MBD_L. This diffusive filling-in reaction is assumed to occur within a syncytium of cell compartments (Fig. 2). Within this cellular syncytium, the feature boundary signals trigger a rapid diffusion, or averaging, of electrical potential across compartment boundaries, except across those compartment boundaries that receive boundary contour signals from the BME stage. The feature contour signals from stage M_L to stage MBD_L are hereby smoothed within stage MBD_L between successive boundary contours that are considered perceptually significant by the BME stage (Cohen and Grossberg, 1983b, 1984; Grossberg, 1983a, 1983b).

Boundary contour signals from stage BME to stage MBD_L are assumed to prevent diffusion across their target compartments by increasing the electrical resistance of the target compartment membranes. It is assumed that the boundary contour signals achieve this increase in cell resistance by acting as inhibitory signals to their target compartments. At the same time that these inhibitory signals increase membrane resistances, they also help to decrease, or hyperpolarize, the target cell potentials. An analogous discussion shows how signals from the BME stage to the MBD_R stage define the boundaries that contain the filling-in reactions which occur in response to feature contour signals from the M_R stage to the MBD_R stage.

5. DOUBLE STEP BRIGHTNESS ILLUSION.

Using these interactions between boundary contour signals, feature contour signals, and diffusive filling-in, a large number of paradoxical brightness percepts have been simulated (Cohen and Grossberg, 1983b, 1984). One well-known phenomenon was studied by Arend, Buehler, and Lockhead (1971) during their investigations of the Craik-O'Brien illusion (O'Brien, 1958).

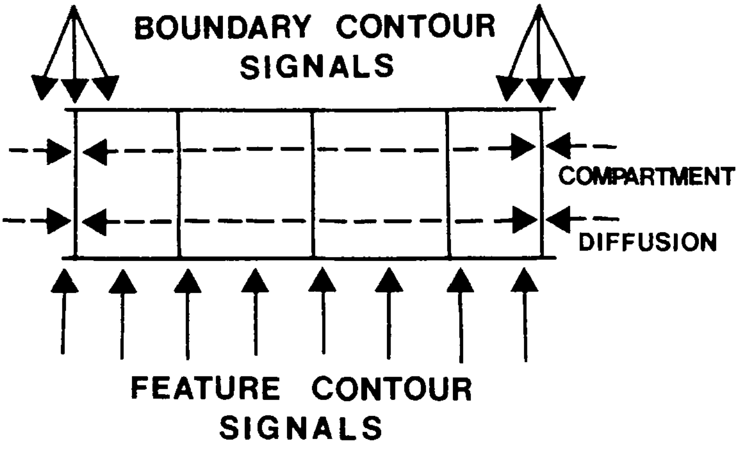

BOUNDARY CONTOUR SIGNALS
COMPARTMENT
DIFFUSION
FEATURE CONTOUR SIGNALS

Figure 2

A monocular brightness domain in stage MBD_L or stage MBD_R. The feature contour signals excite cell compartments which permit rapid lateral diffusion, or filling-in, of activity across their membranes, except at those membranes which receive boundary contour signals from the BME stage. These boundary contour signals inhibit their target membranes and increase their resistance against the lateral flow of activation. The feature contour signals are thus averaged within domains whose boundaries are considered perceptually significant by the BME stage.

Figure 3a describes a luminance profile that was generated by rapidly rotating a disk covered with appropriately cut sectors of black and white papers. The luminances to the left and to the right of the luminance cusps are equal, and the average luminance across the cusps equals the background luminance. Figure 3b shows that this luminance profile is perceived as (approximate) steps of increasing brightness.

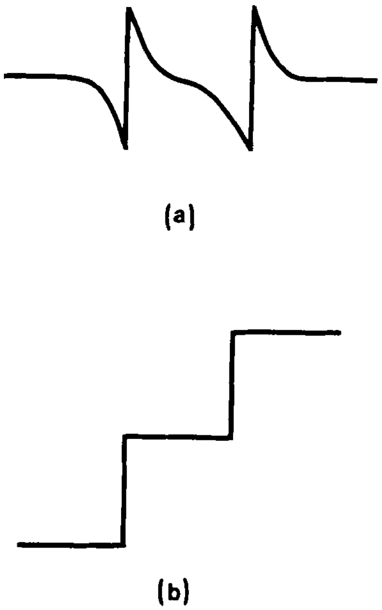


Figure 3

(a) A double Craik-O'Brien luminance profile in which the luminances to the left and right of the luminance cusps are equal, and the average luminance across the cusps equals the background luminance. (b) This luminance profile appears to an observer as two (approximate) steps of brightness.

A similar result is generated by our theory (Figure 4). In the theory, each of the two luminance cusps in the input pattern generates a boundary contour that restricts the lateral diffusion of featural activity. These two boundaries are depicted by the reduced diffusion coefficients in Figure 4. The feature contour process attenuates the background luminance of the input pattern and enhances the contrasts of the cusps. The boundary contour signals also contribute to the total input pattern that is received by the cell compartments. This total input pattern is called the filtered input pattern with boundaries in Figure 4. (The flanks of this total input pattern were artificially extended to the left and to the right to avoid contamination by boundary effects, because our computer could simulate only 2500 cells in a reasonable amount of time). When this total input is allowed to diffuse within the domains defined by the boundary contour signals, the steplike monocular preprocessed pattern of Figure 4 is the result. Appendix A describes the network equations that generate the patterns in Figure 4.

6. NEON COLOR SPREADING AND COMPLEMENTARY COLOR INDUCTION

The phenomenon of neon color spreading illustrates the existence of boundary contours and of feature in a particularly vivid way. Redies and Spillman (1981) have, for example, reported an experiment using a solid red cross and an Ehrenstein figure. When the solid red cross is perceived in isolation, it looks quite uninteresting (Figure 5a). When an Ehrenstein figure is perceived in isolation, it generates an illusory contour whose shape (e.g., circle or diamond) depends upon the viewing distance. When the red cross is placed inside the Ehrenstein figure, the red color flows out of its containing contours and tends to fill the illusory figure (Figure 5b).

In our theory, this percept is explained by showing how boundary contours of the Ehrenstein figure inhibit contiguous boundary contours of the red cross. This inhibitory process takes place within the BME stage as part of the contour completion process (Grossberg and Mingolla, 1984). This inhibitory action within the BME does not prevent the processing of feature contour signals from stage M_L to stage MBD_L and from stage M_R to stage MBD_R. Thus the red contour signals are received by MBD_L and MBD_R despite the fact that some of their corresponding boundary contour signals are inhibited within the BME stage.

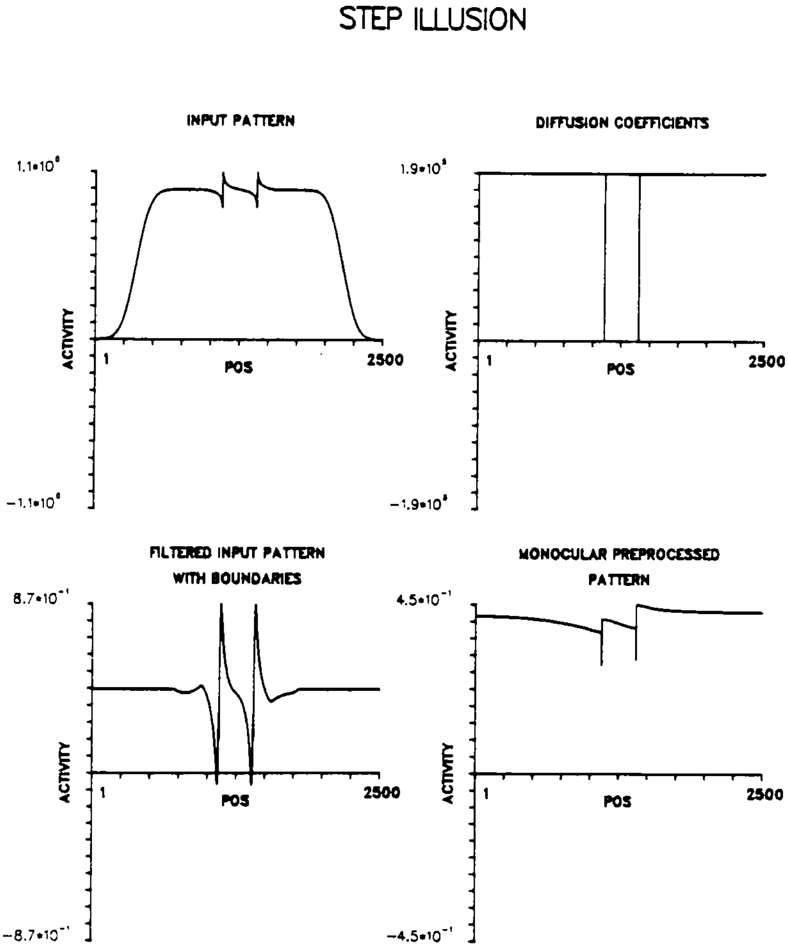


Figure 4

Simulation of the double step brightness illusion. The text describes how the double Craik-O'Brien input pattern generates boundary contours and feature contours that interact to form the step-like output pattern.

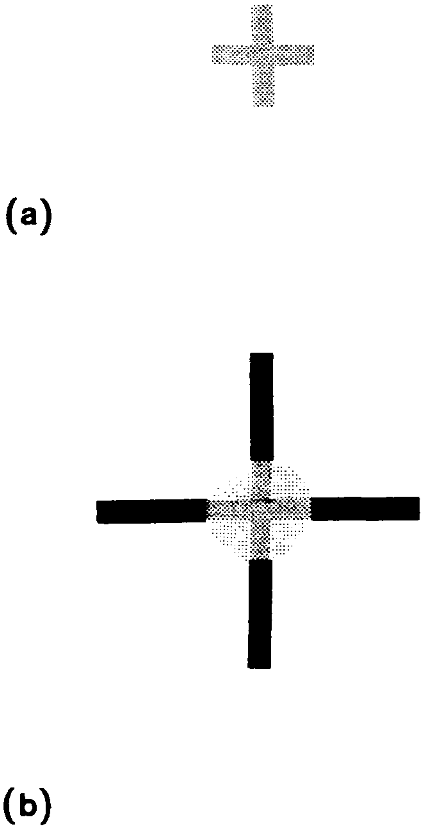


Figure 5

(a) A red cross seems unremarkable when perceived in isolation. (b) When perceived within an Ehrenstein figure, the red color flows outside of the cross and is contained by the illusory contour that is generated by the Ehrenstein figure.

The inhibition of these boundary contour signals within the BME stage allows the red featural activity to diffuse outside of the red cross. The illusory boundary contour that is induced by the Ehrenstein figure restricts the diffusion of this red-labelled activation. This example of neon color spreading thus provides a vivid example in which one can "see" the difference between boundary contours and feature contours, as well as the role of illusory boundary contours in restricting the diffusion of featural activity.

Redies and Spillman (1981) have systematically varied the distance of the red cross from the Ehrenstein contour, their relative orientations, their relative sizes, etc. to study how the strength of the spreading effect changes with scenic parameters. Of particular interest is the fact that "thin [red] flanks running alongside the red connecting lines" (p. 679) can occur if the Ehrenstein figure is slightly separated from the cross. In our theory, separating the Ehrenstein figure from the cross weakens the inhibitory effect of Ehrenstein boundary contours upon the boundary contours of the cross. When the boundary contours of the cross are partially weakened, they can still partially restrict the diffusion of red-labelled activation. Then the red color can only "bleed" outside the contours of the cross.

One might object to this explanation by questioning the ability of the Ehrenstein boundary contours to inhibit the boundary contours of the cross. Why do not the Ehrenstein boundary contours also inhibit contiguous Ehrenstein boundary contours ? If they do, then how do any boundary contours survive this process of mutual inhibition ? If they do not, then doesn't this explanation of neon color spreading collapse ?

At this point, let us recall that the boundary contour process is sensitive to the amount of contrast, even though it is insensitive to the direction of contrast. I claim that the Ehrenstein boundary contours do mutually inhibit one another, but that this inhibition is a type of shunting lateral inhibition such that equally strong inhibitory contour signals can remain positive and balanced (Grossberg, 1973, 1983a). If, however, the Ehrenstein boundary signals are stronger than the boundary signals of the cross by a sufficient amount, then the latter signals can be significantly inhibited. This formal property of the contour completion process provides an explanation of the empirical fact that neon color spreading occurs only if the contrast of a figure (e.g., the cross) relative to the background illumination is less than the contrast of the bounding contours (e.g., the Ehrenstein figure) relative to the background illumination (van Tuijl and de Weert, 1979).

A similar argument helps to explain the complementary color induction that van Tuijl (1975) reported in his original article about the neon effect. Draw on white paper a regular grid of horizontal and vertical black lines that form 5 mm. squares. Replace a subset of black lines by blue lines. Let this subset of lines be replaced, say, from the smallest imaginary diamond shape that includes complete vertical or horizontal line segments of the grid. When an observer inspects this pattern, the blue color of the lines appears to spread around the blue line elements until it reaches the subjective contours of the diamond shape. This percept has the same explanation as the percept in Figure 5b.

Next replace the black lines by blue lines and the blue lines by black lines. Then the illusory diamond looks yellow rather than blue. Let us suppose that the yellow color in the diamond is induced by the blue lines in the background matrix. Then why in the previous display is not a yellow color in the background induced by the blue lines in the diamond ? Why is the complementary color yellow perceived when the background contains blue lines, whereas the original color blue is perceived when the diamond contains blue lines ? What is the reason for this asymmetry ?

I suggest that the reason is as follows. When the diamond is composed of blue lines, then opponent color processing enables the blue lines to induce contiguous yellow contour signals in the background. However these yellow contour signals are constrained by the boundary contour signals of the black lines to remain within the spatial domain of the feature contour of the black lines. Thus the yellow color is not seen in the background. By contrast, the boundary contour signals of the black lines in the background inhibit the contiguous boundary contour signals of the blue lines in the diamond. Hence the blue feature contour signals of the blue lines can flow within the diamond.

When blue lines form the background, they have two effects on the diamond. They induce yellow feature contours via opponent processing. They also inhibit the boundary contour signals of the contiguous black lines. Hence the yellow color can flow within the diamond.

To carry out this explanation quantitatively, we need to study how opponent color processes (red-green, yellow-blue) preprocess the feature contour signals from stage M_L to stage MBD_L and from stage M_R to stage MBD_R. Opponent color processes with the requisite properties are defined by gated dipole fields (Grossberg, 1980). We also need to quantitatively study how the contour completion process responds to complex spatial patterns such as grids and Ehrenstein figures (Grossberg and Mingolla, 1984).

7. FILLING-IN DUE TO STABILIZED IMAGES.

The distinction between boundary contours and feature contours also helps to explain percepts that are induced by stabilized images (Yarbus, 1967). Suppose, for example, that a large circle is surrounded by a uniform red background (Figure 6). Let the circle be divided down the middle by a vertical line. Within each hemicircle, place a small circle filled with the same red color as the background. Color the remainder of the left hemifield black and the remainder of the right hemifield white.

Now stabilize the retinal images of the large circular contour and its vertical edge. The background red is then seen to flow over the black and white hemifields. Since the background red color is the same color as the red color that fills the small circles, one might have supposed that the entire figure would look uniformly red. It does not. The red circle that lies within the previously black hemifield looks brighter, whereas the red circle that lies within the previously white hemifield looks darker.

To explain this percept, let us suppose that stabilizing a contour causes its boundary contour signals to decay (Grossberg, 1983b). Then the red feature contours that are exterior to the large circle can freely flow into the interior of the circle, even as the black and white featural contours of the interior of the circle intermix with one another and the red background. This filling-in process does not, however, enter the small red circles, because their scenic contours are not retinally stabilized.

The relative contrast of the red feature contour within the left red circle is greater than the relative contrast of the red feature contour within the right red circle. These red feature contours are contained by the uninhibited boundary contour signals around each small circle. When the

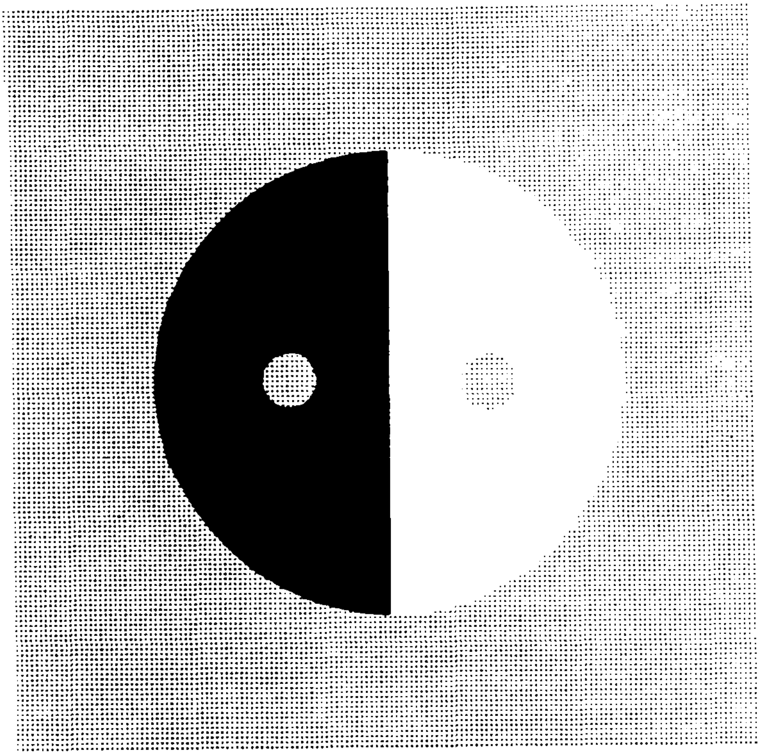


Figure 6

When the boundary of the large circle and the vertical line are stabilized on the retina, the red color outside the large circle fills-in the circle except at the small red circles. The left small circle appears as a brighter red, whereas the right small circle appears as a darker red than the filled-in red color, even though all the red regions are colored using the same red shade.

filling-in process takes place, these relative contrasts are therefore preserved as the total level of featural activity becomes more uniformly distributed around them. The net effect is the obtained percept.

8. THE LAND COLOR AXIOMS.

Land (1977) has developed a formal model to explain the remarkable brightness and color effects that he has demonstrated using McCann Mondrians. This model computes the cumulative relative contrast changes that occur when one serially traverses a path across a picture, and then averages the outcomes across many such paths. For brightness predictions, only shades of gray need to be considered. For color predictions, the computations are done with respect to more than one wavelength. This procedure leads to excellent predictions about the perceived brightnesses and colors of many visual displays. Despite these formal successes, it is difficult to physically interpret Land's serial scanning and serial averaging processes. Moreover, his formal procedure also fails to predict the correct result in a number of imortant cases. A simple example that illustrates the basic nature of these failures is shown in Figure 7.

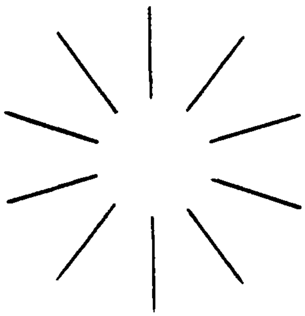


Figure 7

The fact that the illusory circle within the dark lines is brighter than its surround cannot be explained by the Land (1977) retinex theory. The present theory can explain illusory brightness differences and provides a physical interpretation of the formal operations of Land's retinex theory.

In Figure 7, an illusory circle is clearly seen at the inner tips of the radial lines. This illusory circle encloses a region of enhanced illusory brightness. No matter how many radially oriented serial scans are made between the radial lines, their average contrast change cannot predict this illusory brightness difference, because the contrast does not change at all during these scans. Including the dark radial lines within the scans does not improve this incorrect prediction, because the illusory brightness difference does not depend on the number, length, or contrast of these dark lines in a simple way. For example, the strength of the illusory brightness effect is reduced as the angle formed by the lines with the illusory circle decreases (Kennedy, 1978). This happens even if the number, length, and contrast of the lines remain unchanged. This empirical property cannot be explained by any theory that depends only on averages of local contrast changes. The effects are clearly due to the global configuration of the lines. In a similar fashion, the Land (1977) retinex theory cannot explain why rearranging the orientation of the lines in Figure 7 can change the shape of the region where illusory brightness occurs (Kennedy, 1975), or why a single line causes no illusory brightness change whereas even four such lines, if properly configured, can cause a significant illusory brightness change.

The present theory can deal with the illusory effects by showing how the contour completion process can generate illusory contours, and how feature contours can interact with illusory boundary contours to generate illusory brightness changes (Grossberg and Mingolla, 1984). Our theory also clarifies why Land's formal operations often work so well by giving these operations a physical interpretation. Serial computations of contrast changes along multiple paths are replaced by parallel computations of contrast-sensitive feature contours. Averages across serial scans are replaced by diffusive (hence, averaging) filling-in within domains defined by boundary contours. To quantitatively carry out this type of explanation for color percepts, we also need to preprocess the feature contour signals from stage M_L to stage MBD_L and from stage M_R to stage MBD_R using opponent color processes (red-green, yellow-blue) that are defined by gated dipole fields (Grossberg, 1980).

9. FILLING-IN RESONANT EXCHANGE.

All of the processing stages described above are assumed to be pre-perceptual. Only activity in the FIRE stage of Figure 1 is assumed to enter conscious perception. Signals form stage MBD_L and/or MBD_R that are capable of activating the FIRE stage are said to "lift" the preprocessed monocular patterns into the perceptual domain. I assume that only contours of the MBD_L and MBD_R activity patterns input to the FIRE stage. These contours should not be confused with the edges of a visual scene. They are nonlinear statistical constructs of a rather high order by the time they generate signals from the MBD_L stage or the MBD_R stage to the FIRE stage. One needs only to remember that the contour completion process within the BME stage gives rise to the contours within MBD_L and MBD_R to appreciate this fact.

The scale-specific contour signals from the MBD_L and MBD_R stages undergo a binocular matching process in the FIRE stage that is analogous to the matching process whereby boundary contours from the M_L and M_R stages are matched at the BME stage. Only contours that survive this binocular matching process can send feedback signals back to the MBD_L and/or MBD_R stages. These local feedback signals initiate resonant exchanges between the FIRE stage and the MBD stages. These resonant exchanges selectively lift some, but not all, of the preprocessed monocular subdomains into conscious perception.

Once such a filling-in resonant exchange is initiated by a feedback exchange of matched contours, it can rapidly propagate across thousands of receptive fields within its spatial scale. Each such resonant exchange generates a standing wave of patterned activity across a subdomain of the FIRE stage. The ensemble of all such standing waves constitutes the network's percept.

The properties of these filling-in resonant exchanges lie beyond the scope of this chapter. The concept of a FIRE stage was introduced in Grossberg (1981, 1983a) and extensively analysed in Cohen and Grossberg (1983a, 1984).

10. BINOCULAR RIVALRY, STABILIZED IMAGES AND THE GANZFELD.

Three qualitative properties of these resonances are, however, easily noted. These properties further clarify how binocular rivalry and the fading of stabilized images and of a ganzfeld occur within the network of Figure 1.

Suppose that, due to binocular matching, some left-monocular boundary contours are suppressed within the BME stage. Then these boundary contours cannot send boundary contour signals to the corresponding region of stage MBD_L. Featural activity then quickly diffuses across the network positions corresponding to these suppressed contours (Gerrits and Vendrick, 1970). Consequently, no contour signals can be emitted from these positions within the MBD_L stage to the FIRE stage. No edge matches within the FIRE stage can occur at these positions, and no feedback signals are returned from the FIRE stage to the MBD_L stage to lift the corresponding monocular subdomain into perception. Thus the subdomains whose boundary contours are suppressed within the BME stage are no longer perceived. As soon as the corresponding boundary contours win the BME binocular competition, however, these subdomains can again rapidly support the resonant lifting of their subdomain into perception at the FIRE stage.

A similar argument holds for the nonperception of a subdomain whether its boundary edges are suppressed by binocular rivalry within the BME stage, by image stabilization, or do not exist, as in a ganzfeld.

REFERENCES.

[1] Arend, L.E., Buehler, J.N., and Lockhead, G.R., Difference information in brightness perception, Perception and Psychophysics, 9 (1971), 367-370.

[2] Blake, R, and Fox, R., Binocular rivalry suppression: insensitive to spatial frequency and orientation change, Vision Research, 14 (1974) 687-692.

[3] Carpenter, G.A., and Grossberg, S., Adaption and transmitter gating in vertebrate photoreceptors, Journal of Theoretical Neurobiology, 1 (1981), 1-42.

[4] Cogan, A.I., Monocular sensitivity during binocular viewing, Vision Research, 22 (1982), 1-16.

[5] Cohen, M.A., and Grossberg, S., Some global properties of binocular resonances: disparity matching, filling-in, and figure-ground synthesis, in P. Dodwell and T. Caelli (Eds.), Figural synthesis, Hillsdale, NJ., Erlbaum Associates (1983a).

[6] Cohen, M.A., and Grossberg, S., Neural dynamics of binocular form perception, Society for Neuroscience Abstracts, 13th Annual Meeting, Boston, 353.8 (1983b).

[7] Cohen, M.A., and Grossberg, S., The dynamics of brightness perception: a psychological and a neural theory, Submitted for publication (1984).

[8] Gellatly, A.R.H., Perception of an illusory triangle with masked inducing figure, Perception, 9 (1980), 599-602.

[9] Gerrits, H.J.M., de Haan, B., and Vendrick, A.J.H., Experiments with retinal stabilized images: relations between the observations and neural data, Vision Research, 6 (1966), 427-440.

[10] Gerrits, H.J.M., and Timmerman, J.G.M.E.N., The filling-in process in patients with retinal scotomata, Vision Research, 9 (1969), 439-442.

[11] Gerrits, H.J.M., and Vendrick, A.J.H., Simultaneous contrast, filling-in process and information processing in man's visual system, Experimental Brain Research, 11 (1970), 411-430.

[12] Grossberg, S., Contour enhancement, short-term memory, and constancies in reverberating neural networks, Studies in Applied Mathematics, 52 (1973), 217-257.

[13] Grossberg, S., How does a brain build a cognitive code?, Psychological Review, 87 (1980), 1-51.

[14] Grossberg, S., Adaptive resonance in development, perception, and cognition, in S. Grossberg (Ed.), Mathematical psychology and psychophysiology, Providence, RI, American Mathematical Society (1981).

[15] Grossberg, S., Studies of mind and brain: neural principles of learning, perception, development, cognition, and motor control, Boston, Reidel Press (1982).

[16] Grossberg, S., The quantized geometry of visual space: the coherent computation of depth, form, and lightness, The Behavioral and Brain Sciences, in press (1983a).

[17] Grossberg, S., Neural substrates of binocular form perception: filtering, matching, diffusion, and resonance, in E. Basar, H. Flohr, H. Haken, and A.J. Mandell (Eds.), Synergetics of the brain, New York, Springer Verlag (1983b).

[18] Grossberg, S., and Mingolla, E., Neural dynamics of form perception: illusory figures and neon color spreading, submitted for publication (1984).

[19] Hendrickson, A.E., Hunt, S.P., and Wu, J.-Y., Immunocytochemical localization of glutamic acid decarboxylase in monkey striate cortex, Nature,

292 (1981), 605-607.

[20] Horton, J.C., and Hubel, D.H., Regular patchy distribution of cytochrome oxidase staining in primary visual cortex of macaque monkey, Nature, 292 (1981), 762-764.

[21] Hubel , D.H., and Livingstone, M.S., Regions of poor orientation tuning coincide with patches of cytochrome oxidase staining in monkey striate cortex, Society for Neuroscience Abstract, 11th Annual Meeting, Los Angeles, 118.12 (1981).

[22] Hubel, D.H., and Wiesel, T.N., Functional architecture of macaque monkey visual cortex, Proceedings of the Royal Society of London (B), 198 (1977), 1-59.

[23] Julesz, B., Foundations of cyclopean perception, Chicago, University of Chicago Press (1971).

[24] Kanizsa, G., Contours without gradients or cognitive contours?, Italian Journal of Psychology, 1 (1974), 93-113.

[25] Kaufman, L., Sight and mind: an introduction to visual perception, New York, Oxford University Press (1974).

[26] Kennedy, J.M., Depth at an edge, coplanarity, slant depth, change in direction, and change in brightness in the production of subjective contours, Italian Journal of Psychology, 2 (1975), 107-123.

[27] Kennedy, J.M., Illusory contours and the ends of lines, Perception, 7 (1978), 605-607.

[28] Kennedy, J.M., Subjective contours, contrast, and assimilation, in C.F. Nodine, and D.F. Fisher (Eds.), Perception and pictorial representation, New York, Praeger (1979).

[29] Kennedy, J.M., Illusory brightness and the ends of petals: change in brightness without aid of stratification or assimilation effects, Perception, 10 (1981), 583-585.

[30] Kulikowski, J.J., Limit of single vision in stereopsis depends on contour sharpness, Nature, 275 (1978), 126-127.

[31] Land, E.H., The retinex theory of color vision, Scientific American, 237 (1977), 108-128.

[32] Livingstone, M.S., and Hubel, D.H., Thalamic inputs to cytochrome oxidase-rich regions in monkey visual cortex, Proceedings of the National Academy of Sciences, 79 (1982), 6098-6101.

[33] Marr, D., and Hildreth, E., Theory of edge detection, Proceedings of the Royal Society of London (B), 207 (1980), 187-217.

[34] O'Brien, V., Contour perception, illusion, and reality, Journal of the Optical Society of America, 48 (1958), 112-119.

[35] Parks, T.E., Subjective figures: some unusual concomitant brightness effects, Perception, 9 (1980), 239-241.

[36] Parks, T.E., and Marks, W., Sharp-edges vs. diffuse illusory circles: the effects of varying luminance, Perception and Psychophysics, 33 (1983), 172-176.

[37] Petry, S., Harbeck, A., Conway, J., and Levey, J., Stimulus determinants of brightness and distinctness of subjective contours, Perception and Psychophysics, 34 (1983), 169-174.

[38] Pritchard, R.M., Stabilized images on the retina, Scientific American, 204 (1961), 72-78.

[39] Pritchard, R.M., Heron, W., and Hebb, D.O., Visual perception approached by the method of stabilized images, Canadian Journal of Psychology, 14 (1960), 67-77.

[40] Redies, C., and Spillman, L. The neon color effect in the Ehrenstein illusion, Perception, 10 (1981), 667-681.

[41] Riggs, L.A., Ratliff, F., Cornsweet, J.C., and Cornsweet, T.N., The disappearance of steadily fixated visual test objects, Journal of the Optical Society of America, 43 (1953), 495-501.

[42] van Tuijl, H.F.J.M., A new visual illusion: neonlike color spreading and complementary color induction between subjective contours, Acta Psychologica, 39 (1975), 441-445.

[43] van Tuijl, H.F.J.M., and de Weert, C.M.M., Sensory conditions for the occurrence of the neon spreading illusion, Perception, 8 (1979), 211-215.

[44] van Tuijl, H.F.J.M., and Leeuwenberg, E.L.J., Neon color spreading and structural information measures, Perception and Psychophysics, 25 (1979), 269-284.

[45] Yarbus, A.L., Eye movements and vision, New York, Plenum Press (1967).

APPENDIX

This appendix describes the neural network that Cohen and Grossberg (1984) use to simulate brightness percepts. The following simulations were done on one-dimensional fields of cells. The input pattern $(I_1, I_2, \ldots, I_n)$ is transformed into the output pattern $(z_1, z_2, \ldots, z_n)$ via the following equations.

FEATURE CONTOURS

The input pattern $(I_1, I_2, \ldots, I_n)$ is transformed into feature contours via a feedforward on-center off-surround network of cells undergoing shunting, or membrane equation, interactions. The activity, or potential x_i, of the i^{th} cell in a feature contour pattern is

$$\frac{d}{dt} x_i = -Ax_i + (B - x_i)\sum_{k=1}^{n} I_k C_{ki} - (x_i + D)\sum_{k=1}^{n} I_k E_{ki} . \tag{1}$$

Both the on-center coefficients C_{ki} and the off-surround coefficients E_{ki} are Gaussian functions of intercellular distance; viz., $C_{ki} = C\exp[-\mu(i-k)^2]$ and $E_{ki} = E\exp[-\nu(i-k)^2]$, where $\mu > \nu$. System (1) is assumed to react much more quickly than the diffusive filling-in process. Hence, we assume that each x_i is in approximate equilibrium with respect to the input pattern. At equilibrium, $\frac{d}{dt} x_i = 0$ and

$$x_i = \frac{\sum_{k=1}^{n}(BC_{ki} - DE_{ki})I_k}{A + \sum_{k=1}^{n}(C_{ki} + E_{ki})I_k} . \tag{2}$$

The activity pattern $(x_1, x_2, \ldots, x_n)$ is sensitive to both the amount and the direction of contrast in edges of the input pattern (Grossberg (1983a)). These feature contour activities generate inputs of the form

$$F_i = \frac{x_i}{1 + \alpha S_i}$$

to the diffusive filling-in process. The inhibitory term S_i is defined by the boundary contour process in equation (7) below.

BOUNDARY CONTOURS

The input pattern $(I_1, I_2, \ldots, I_n)$ also activates the boundary contour process, which we represent as a feedforward on-center off-surround networking undergoing shunting interactions. This simplified view of the boundary contour process is permissible in the present simulations because the simulations, being one-dimensional and monocular, do not need to account for orientational tuning or binocular matching. Since the simulations do not probe the dynamics of illusory contour formation, the contour completion process can also be ignored.

As in equation (2), the input pattern rapidly gives rise to an activity pattern

$$y_i = \frac{\sum_{k=1}^{n} (\tilde{B}\tilde{C}_{ki} - \tilde{D}\tilde{E}_{ki})I_k}{\tilde{A} + \sum_{k=1}^{n} (\tilde{C}_{ki} + \tilde{E}_{ki})I_k} \tag{4}$$

where $\tilde{C}_{ki} = \tilde{C} \exp[-\tilde{\mu}(i-k)^2]$ and $\tilde{E}_{ki} = \tilde{E} \exp[-\tilde{\nu}(i-k)^2]$. The Gaussian coefficients $\tilde{\mu}$ and $\tilde{\nu}$ are, however, larger than the Gaussian coefficients μ and ν, because boundary contours are narrower than feature contours.

The activity pattern $(y_1, y_2, \ldots, y_n)$ is sensitive both to the direction and amount of contrast in the input pattern $(I_1, I_2, \ldots, I_n)$. The sensitivity to the direction of contrast is progressively eliminated by the following operations. Let the output signal from BME to MBD that is elicited by activity y_i equal $f(y_i)$, where $f(w)$ is a sigmoid signal of the rectified part of y_i; viz.,

$$f(y_i) = \frac{\beta\{[y_i]^+\}^\gamma}{1 + \delta\{[y_i]^+\}^\gamma} . \tag{6}$$

The notation $[w]^+ = \max(w,0)$ and $\gamma > 1$. The output signals $f(y_i)$ are spatially distributed before influencing cell compartments of the cell syncytium. The total signal to the i^{th} cell compartment due to the activity pattern $(y_1, y_2, \ldots, y_n)$ is

$$S_i = \sum_{k=1}^{n} G_{ik} f(y_k), \tag{7}$$

where G_{ik} is a Gaussian function of intercellular distance; viz., $G_{ik} = G \exp[-\omega(i-k)^2]$. This Gaussian fall-off (ω) is less narrow than the boundary contour process $(\tilde{\mu},\tilde{\nu})$, but more narrow than the feature contour process (μ,ν).

DIFFUSIVE FILLING-IN

The activity z_i of the i^{th} cellular compartment of the cellular syncytium obeys the nonlinear diffusion equation

$$\frac{d}{dt} z_i = -Hz_i + J_{i+1,i}(z_{i+1} - z_i) + J_{i-1,i}(z_{i-1} - z_i) + F_i, \tag{8}$$

where the input F_i is defined by equation (3). The diffusion coefficients $J_{i+1,i}$ and $J_{i-1,i}$ are determined by boundary contour signals according to equations of the form

$$J_{i+1,i} = \frac{\lambda}{1 + \kappa[S_{i+1} - \Gamma]^+ + \kappa[S_i - \Gamma]^+} \tag{9}$$

and

$$J_{i-1,i} = \frac{\lambda}{1 + \kappa[S_{i-1} - \Gamma]^+ + \kappa[S_i - \Gamma]^+}, \tag{10}$$

where the threshold $\Gamma > 0$. Thus an increase in the boundary signal S_i decreases both diffusion coefficients $J_{i+1,i}$ and $J_{i-1,i}$. The feature contour signal F_i also decreases when the boundary signal S_i increases. In all the equations (3), (9), and (10), the inhibitory effects of boundary signals S_i on cell compartment membranes act via shunting inhibition. A positive threshold Γ occurs in (9) and (10), but not in (3), because we assume that the intercompartmental membranes that regulate diffusion of activity between compartments are less accessible to the signals S_i than are the exterior surface membranes that bound the cellular syncytium.

Supported in part by the Air Force Office of Scientific Research (AFOSR 82-0148).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

ATTENTION IN DETECTION THEORY

H. C. Micko
Institute of Psychology
University of Technology
Brunswick, Germany

Detection theory takes account of the conditional hit and false alarm probabilities, given stimuli have been inspected and decided upon. From that exact formulation a theory is derived which permits, under adequate experimental conditions, the separate estimation of attention, discrimination and bias parameters.

1. DEFINITIONS

Detection theory (see e.g. Green, and Swets (1966)) is expressed in most general form by the pair of equations

$$p(Y|s) = f_s(d,b,\ldots), \quad p(Y|n) = f_n(d,b,\ldots), \tag{1}$$

with s denoting the event of a signal, n that of a nonsignal or noise stimulus being presented to an observer O. Y denotes the event of O reporting the stimulus, s or n, as a signal. In other words, $p(Y|s)$ and $p(Y|n)$ denote hit and false alarm probabilities to be estimated by empirical hit and false alarm rates. $p(Y|s)$ and $p(Y|n)$ are related by functions f_s resp. f_n to some discrimination parameter, d, and some decision bias parameter, b. Sometimes additional parameters are considered, e.g. signal to noise variance ratio or criterion variance. We shall leave them out of account in order to simplify arguments and notation.

In subsequent equations we shall consider an observation event, i.e. the event of O making a perceptual decision about whether or not to report a signal. We designate that event by o Finally, we let $\bar{Y}$ and $\bar{o}$ denote the events complementary to Y and o, i.e. the events of O not reporting resp. not observing the stimulus presented. Nonobservation may be due to the fact that the stimulus does not arrive at O's perceptual system or that it is not processed up to the response selection stage.

Specific detection, choice and threshold models differ by the pair of functions f_s, f_n postulated, which may reflect different numerical representations and/or interpretations of discrimination and bias. For convenience we define f_s and f_n to be nondecreasing functions of b, larger values of which represent stronger bias towards reporting stimuli as signals. We assume $f_s \geqslant f_n$ for any b and let greater values of d be associated with larger excess of f_s over f_n for any particular b (see Figure 1). Thus, e.g., b is positively related to guessing probability in high threshold models, negatively to beta or C in classical detection models that postulate Gaussian distributions for f_s, f_n; d is positively related to d' in Gaussian detection models and $-\log \eta$ in Luce's (1963) choice model but negatively to η. When considering Gaussian detection models we denote by $_{-\infty}^{\;b}N(0,1)$ the value of the standard normal distribution function at b.

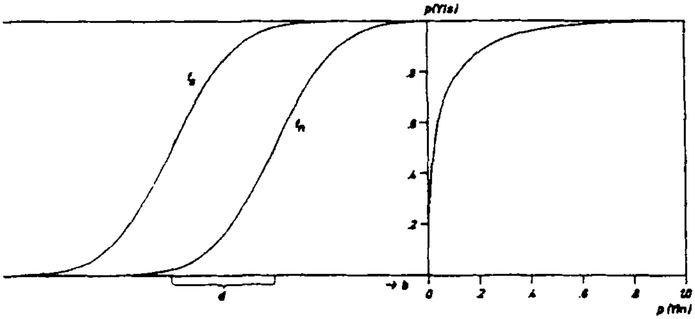


Figure 1

Classical (Gaussian) detection theory.

Left: Decision functions (report probability for signals (f_s) and noise (f_n) as function of bias (b)). Discriminability, $d=2\sigma$.

Right: ROC-Curve of the unconditional Gaussian detection model, $d=2\sigma$.

2. A TACIT ASSUMPTION OF DETECTION THEORY

The usual interpretation of the parameters d and b rests heavily on an implicit but fundamental assumption of detection theory: detection is viewed as a perceptual decision task, i.e. responses Y and $\bar{Y}$ are assumed to reflect the outcome of a decision about a perceptual event. This assumption is required in order to justify inferences about O's perceptual discrimination and decision bias from his hit and false alarm rates. We make that assumption explicit by rewriting the fundamental equations (1) of detection theory:

$$p(Y|o \wedge s) = f_s(d,b), \quad p(Y|o \wedge n) = f_n(d,b). \qquad (2)$$

Eqs. (2) state, more precisely than eqs. (1), that detection theory takes account of the probabilities of a hit and false alarm, given the inspection of and decision about the stimulus presented. In fact, hits and false alarms may result from arbitrary, perception-independent guesses in the absence of any inspection of the stimulus to be judged, apparent misses and correct rejections may reflect just lack of observation. In both cases the responses Y and $\bar{Y}$ are not outcomes of a perceptual decision and, therefore, cannot properly be taken account of by classical detection theory. As a consequence, hit and false alarm rates that are affected by arbitrary guessing or lack of attention are unsuitable for analysis by detection theory and should be precluded if misleading parameter estimates are to be avoided (it might be mentioned that perception-independent guessing differs from the kind of guessing assumed in threshold models, the probability of which explicitely depends on O's perceptual state "Detect" resp. "Non-detect").

In classical detection experiments it was assumed that all stimuli presented were in fact inspected and judged as required by the theory. That assumption is not unreasonable since O's were informed about time and place of occurrence of the stimuli. We may also hope that sufficient care has been taken to secure the high morale necessary to immunize observers, who made thousands of judgements, against the temptation to produce responses without bothering about the stimuli. The above assumption, however, cannot be upheld, if detection theory is applied to data of attention experiments, because it is the very purpose of attention paradigms to investigate behavior under conditions that make stimuli likely to escape O's notice. It has been recognized that b, when estimated from eqs. (1) instead of eqs. (2), measures the confounded effects of attention and judgement bias (Jerison, Pickett & Stenson (1965); Swets & Kristoffersen (1970); Broadbent (1971), p. 93f.). Similarly, perception-independent guessing may reduce d without the perceptual system being impaired or stimulus discriminability being reduced. Perception-independent guessing, however, may occur in attention experiments if no special precautions are taken. It is bound to occur if O's are forced to respond to every stimulus presented, whether observed or unnoticed. It is the purpose of the present paper to derive a more appropriate formulation of detection theory from eqs. (2) which permits, under adequate experimental conditions, the seperate estimation of attention, discrimination and decision parameters.

3. THEORY OF CONDITIONAL DETECTION

The conditional hit and false alarm probabilities given in eqs. (2) are the appropriate probabilities from which to determine discrimination and bias parameters. As a rule, however, the respective conditional hit and false alarm rates cannot be obtained experimentally, because without special questioning of O's we do not know whether or not a response, Y resp. $\bar{Y}$, resulted from a perceptual decision. In most cases, the available data will be the unconditional hit and false alarm rates which estimate $p(Y|s)$ and $p(Y|n)$. Therefore eqs. (2) is of little practical value so far and we have to derive appropriate expressions for the unconditional probabilities.

We assume that the set of correct or false reports of a signal can be subdivided without remainder or overlap into a set of reports based on perceptual decisions (although possibly derived from weak or faulty evidence) and a set of reports based on perception-independent guessing, i.e.

$$\{Y \wedge s\} = \{Y \wedge o \wedge s\} \cup \{Y \wedge \bar{o} \wedge s\}$$
$$\{Y \wedge n\} = \{Y \wedge o \wedge n\} \cup \{Y \wedge \bar{o} \wedge n\}.$$

From that assumption we obtain for the respective probabilities of occurrence:

$$p(Y \wedge s) = p(Y \wedge o \wedge s) + p(Y \wedge \bar{o} \wedge s)$$
$$p(Y \wedge n) = p(Y \wedge o \wedge n) + p(Y \wedge \bar{o} \wedge n),$$

and dividing by $p(s)$, resp. $p(n)$, yields the following expression for the unconditional hit and false alarm probabilities:

$$\begin{aligned} p(Y|s) &= p(Y \wedge o|s) + p(Y \wedge \bar{o}|s), \\ p(Y|n) &= p(Y \wedge o|n) + p(Y \wedge \bar{o}|n). \end{aligned} \tag{3}$$

Applying the elementary theorem on conditional probabilities, $P(A \wedge B|C) = p(B|C) \cdot p(A|B \wedge C)$, to the probabilities on the right sides of eqs. (3) we obtain

$$\begin{aligned} p(Y|s) &= p(o|s) \cdot p(Y|o \wedge s) + p(\bar{o}|s) \cdot p(Y|\bar{o} \wedge s), \\ p(Y|n) &= p(o|n) \cdot p(Y|o \wedge n) + p(\bar{o}|n) \cdot p(Y|\bar{o} \wedge n). \end{aligned} \tag{4}$$

Equivalent equations were discussed by Wykoff (1963) in the context of discrimination learning.

In eqs. (4) all probabilities which are of relevance to detection and attention do occur. $p(Y|o \wedge s)$ and $p(Y|o \wedge n)$ are the familiar conditional probabilities of eqs. (2) which are taken account of by classical detection theory.

$P(Y|\bar{o} \wedge s)$ and $p(Y|\bar{o} \wedge n)$, are the conditional hit and false alarm probabilities, given the stimulus has not been observed. They are the probabilities of reporting a signal on the basis of an arbitrary guess. Since no perception is involved, the report of a signal is as likely as that of a noise stimulus and the probabilities are obviously independent of d. They may be a function of b, however, because under conditions that make O more likely to judge a perceived stimulus as a signal he might also be more likely to guess that an unobserved stimulus is a signal. Hence we assume

$$p(Y|\bar{o} \wedge s) = p(Y|\bar{o} \wedge n) = f_g(b), \tag{5}$$

with $f_g(b)$ denoting some nondecreasing function of b. The assumption does not preclude the special case $f_g(b) \equiv$ const., or even $f_g(b) \equiv 0$.

$p(o|s)$ and $p(o|n)$ are observation probabilities, i.e. the probabilities of O noticing and deciding upon the stimulus presented. $p(\bar{o}|s) = 1-p(o|s)$ and $p(\bar{o}|n) = 1-p(o|n)$ are the respective complements, i.e. the probabilities of overlooking the stimulus or losing it before it is processed up to the response selection stage. These are the probabilities of interest to the scholar of attention. We may treat them as parameters of the theory or derive them from a separate model, following Jerison, Pickett & Stenson (1967) who argued that detection theory should be applied to observation probabilities rather than to judgement probabilities. Inspection, o, and non-inspection, $\bar{o}$, of a stimulus presented may be regarded as outcomes of a decision just as a report, Y, and non-report, $\bar{Y}$. Hence, we assume, in analogy to eqs. (1)

$$p(o|s) = f'_s(c,a), \quad p(o|n) = f'_n(c,a) \tag{6}$$

with functions f'_s and f'_n not necessarily equal to, but having the same properties as f_s and f_n (see Definitions), c and a as parameters equivalent to d and b. The parameter c may be interpreted as a measure of relative conspicuity of s over n, i.e. as a measure of preattentional discriminability. The parameter a is a bias parameter like b. However, it measures attention bias towards the stimuli presented - as opposed to decision bias.

By f'_s and f'_n observation probability increases as attention towards the stimulus source increases and $p(o|s)$ exceeds $p(o|n)$ progressively with increasing signal conspicuity.

Inserting eqs. (2), (5) and (6) into eqs. (4) we obtain the general equations of a theory of conditional detection:

$$\begin{aligned} p(Y|s) &= f^*_s(d,b,c,a) = f'_s(c,a) \cdot f_s(d,b) + (1-f'_s(c,a)) \cdot f_g(b) \\ p(Y|n) &= f^*_n(d,b,c,a) = f'_n(c,a) \cdot f_n(d,b) + (1-f'_n(c,a)) \cdot f_g(b). \end{aligned} \qquad (7)$$

Eqs.(7) turn out to represent a two step detection theory supplemented by correction for guessing. The first step concerns the decision whether or not inspect the stimulus, the second step concerns the decision of whether or not report it. Classical detection theory (eqs. 1) is a special case obtained by letting $a = \infty$ or equivalently $p(o|s) = p(o|n) = 1$ (see eqs. 7a).

Conditional detection theory may be regarded as an identifiable and testable formulation of ideas advanced by Jerison, Pickert & Stenson (1965) who postulated three states of attention "alert", "blurred" and "distracted", each of which ought to be taken account of by a detection model of its own. Their "distracted" state is about equivalent to our non-observation state $\bar{o}$. "Alertness" vs "blurredness" is treated here as a continuous variable to be measured by the sensitivity parameter d.

In subsequent sections we shall express eqs. (7) also in the less explicit form

$$\begin{aligned} p(Y|s) &= f^*(d,b,p(o|s)) = p(o|s) \cdot f(d,b) + (1-p(o|s)) \cdot f_g(b), \\ p(Y|n) &= f^*(d,b,p(o|n)) = p(o|n) \cdot f(d,b) + (1-p(o|n)) \cdot f_g(b), \end{aligned} \qquad (7a)$$

in order to make arguments more apprehensible. In eqs. (7a) the observation probabilities are treated as parameters which may or may not be further explained by an attention model of the type suggested above (eqs. 6).

In order to apply the theory to data, three functions, f'_s, f'_n and f'_g, have to be specified in addition to f_s and f_n, with $f'_s = f_s$, $f'_n = f_n$ being an obvious candidate and a high threshold model $f'_s = c + (1-c)a$, $f'_n = a$, another. Two parameters, c and a, have to be estimated in addition to the traditional parameters d, b etc. Therefore, parameter estimation and goodness of fit testing require the empirical determination of several points of the ROC-curve. That may be awesome but surely is of educational value, because it prevents the simple computation of d and b from a single pair of hit and false alarm rates without applying a goodness of fit test to the model employed. The reduction of parsimony in comparison with unconditional models may be accepted in view of the increased precision of interpretation and the avoidance of confounding attentional effects on hit and false alarm rates with perceptional and judgemental effects. Moreover, not much parsimony can be lost, considering that almost any empirical ROC-curve can be fitted by one of the many unconditional models, f_s, f_n, already proposed (Laming, (1973)).

The interpretation of conditional detection theory and its parameters is testable too, since attention and judgement bias can be manipulated independently. Any test, admittedly, rests on the assumption that effects of attention-, discrimination- and decision-variables on hit and false alarm rates do not interact. It may be predicted, e.g., that confidence ratings represent decision bias and do not vary with shifts of attention. On the other hand, distractions ought to reduce attention bias but should not influence decision criteria. Both kinds of bias should also be affected independently by appropriate manipulations of the pay-off matrix, decision bias by varying the ratio of pay-offs for reports versus rejections, attention bias by proportional increases of all four entries of the pay-off matrix. The separate manipulation of conspicuity and discriminability is more difficult. In a discrimination paradigm the difference between the two classes of stimuli to be discriminated may be manipulated on a variable which is unrelated to conspicuity. E.g., s and n may be tones of different frequency but equal intensity. Thus, d varies with frequency difference whereas $c = 0$ remains constant. In signal detection and attention paradigms, however, c and d will usually be confounded. In that case, $c = d$ may be a reasonable a priori assumption. In any case, we have to assume $c \leq d$, because we cannot expect preattentional discrimination to be superior to overall discriminability. If noise stimuli are more conspicuous than signals, c may take negative values, resulting in $p(o|s) < p(o|n)$. By assuming $c = 0$ or $c = d$, the number of parameters of a conditional detection model may be reduced by one to a minimum of three.

4. MODELS WITH NON-CONSTANT GUESSING FUNCTIONS

The form of an ROC-curve is determined by the detection model f_s, f_n, adopted. That of an ROC-curve predicted from a model of conditional detection depends heavily, also, on the guessing function f_g assumed - particularly with respect to those properties by which the appropriateness of a conditional model can be evaluated. The probability $p(Y|\bar{o}\wedge s) = p(Y|\bar{o}\wedge n)$ of reporting a stimulus on the basis of a pure guess may or may not co-vary with O's decision bias b. In the former case $f_g(b)$ is a monotoneously increasing function of b, in the latter $f_g(b) \equiv$ const. In this section we consider the former case.

Many, if not most detection models, can be written in the form

$$f_s(d,b,m) = f(b+m\cdot d), \quad f_n(d,b,m) = f(b-(1-m)\cdot d), \quad 0 \leq m \leq 1, \tag{8}$$

i.e. f_s and f_n differ only by a translation from each other on the (appropriately transformed) decision axis. Equivalently, we may say, f_s and f_n differ by a translation from some reference function $f_r(b)$ which is situated somewhere between f_s and f_n (e.g. Gaussian detection theory: $f_r = {}_{-\infty}^{b}N(0,1)$, $f_s = {}_{-\infty}^{b+m\cdot d}N(0,1)$, $f_n = {}_{-\infty}^{b-(1-r)\cdot d}N(0,1)$; logistic choice theory:

$f_r = \exp(b)/1+\exp(b)$, $f_s = \exp(b+m\cdot d)/1+\exp(b+m\cdot d)$,

$f_n = \exp(b-(1-m)\cdot d)/1+\exp(b-(1-m)\cdot d)$; exponential detection theory:
$f_r = 1/\exp(k^{b})$, $f_s = 1/\exp(k^{b+m\cdot d})$, $f_n = 1/\exp(k^{b-(1-m)d})$.
The distance between f_s and f_n is $m\cdot d+(1-m)\cdot d = d$. The additional parameter m is not a parameter of the model f_s, f_n, but determines the respective distances of f_s and f_n from the so far arbitrarily chosen reference function f_r. For simplicity of presentation, we disregard models that predict functions f_s, f_n of different steepness. The reader will realize, however, that the following arguments and conclusions hold more or less for all models with similar and overlapping functions f_s, f_n (see Figure 2).

If f_s and f_n differ by translation only, it is most natural to assume that the guessing function f_g, too, differs from f_s and f_n by translation only, and that it is located somewhere between f_s and f_n. We let

$$f_g(b) = f_r(b) = f(b), \tag{9}$$

thereby giving the reference function $f_r(b)$ an empirical meaning. Now the parameter m represents the relative closeness of f_g to f_s and f_n and may be assumed to equal a priori signal probability.

Inserting eqs. (8) and (9) in eqs. (7a), we obtain the general expression for conditional detection models with functions $f_s = f_n = f_g = f$ equal up to translation.

$$\begin{aligned} p(Y|s) &= f_s^*(d,b,m,p(o|s)) = p(o|s)\cdot f(b+m\cdot d)+(1-p(o|s))\cdot f(b), \\ p(Y|n) &= f_n^*(d,b,m,p(o|n)) = p(o|n)\cdot f(b+(1-m)\cdot d)+(1-p(o|n))\cdot f(b). \end{aligned} \tag{10}$$

We want to compare these conditional models with the following unconditional detection models:

$$\begin{aligned} p(Y|s) &= f_s^o(d^o,b) = f(b+p(o|s)\cdot m\cdot d), \\ p(Y|n) &= f_n^o(d^o,b) = f(b-p(o|n)\cdot(1-m)\cdot d). \end{aligned} \tag{11}$$

For every conditional model of the form (10) there exists an unconditional model of the form (11) with discriminability parameter

$$d^o = p(o|s)\cdot m\cdot d + p(o|n)\cdot(1-m)\cdot d \leq d \ (!!). \tag{12}$$

f_s^* and f_n^* are obtained by computing for each argument b the weighted mean of $f_s(b)$ and $f_g(b)$, resp. $f_g(b)$ and $f_n(b)$, with observation and guessing probabilities as weights. f_s^o and f_n^o are obtained by translation from f_g, i.e. more precisely by, taking for each pair of equal values of the functions f_s and f_g, resp. f_g and f_n, the weighted mean of their respective arguments (see Figure 2).

Thus both, f_s^* and f_s^o, are weighted averages of f_s and f_g, f_n^* and f_n^o averages of f_g and f_n. As a rule, $f_s^* \neq f_s^o$ and $f_n^* \neq f_n^o$ because averaging across ordinates and across abscissas does not yield the same function. However, f_s^* as well as f_n^o is bounded, over the whole range of b, by f_s from above and by f_g from below, both approach f_s for $p(o|s)\to 1$ and f_g for $p(o|s)\to 0$. In view of the large overlap of f_s and f_n in usual detection situations and, as a consequence, the still larger overlap of f_s and f_g, we conclude that, f_s^* differs only a little from f_s^o. By the same argument, f_n^* and f_n^o are similar functions. Consequently the conditional detection model f_s^*, f_n^* with discriminability parameter d differs only a little from the unconditional detection model f_s^o, f_n^o with the discriminability parameter $d^o \leq d$ given in eq. (12). Thus, if f_g is of the same, or similar, form as f_s and f_n, nonobservation combined with guessing has an effect similar to that of reduced discriminability. Considering the comparatively large sampling errors of empirical hit and false alarm rates, we conclude that the ROC-curve of a conditional detection model will be practically indistinguishable from the ROC-curve of some classical detection model with smaller discrimination parameter.

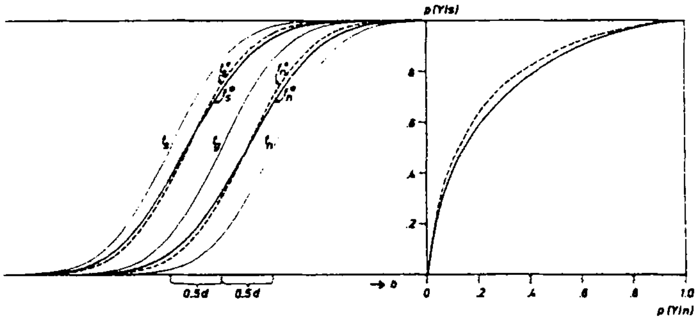


Figure 2

Conditional detection theory with Gaussian decision and guessing functions.

Left: (Solid lines) Report probability for signal (f_s^*) and noise (f_n^*) as weighted mean of the Gaussian decision function f_s resp. f_n of Figure 1 and a Gaussian guessing function f_g equidistant from f_s and f_n.

$f_s^*(b) = .67\ f_s(b)+.33\ f_g(b)$, $f_n^*(b) = .5\ f_n(b)+.5\ f_g(b)$.

(Broken lines) Decision functions of an unconditional Gaussian detection model (f_s^o, f_n^o) with discriminability parameter $d^o = 1.16$ ($= .58 \cdot d = .67 \cdot .5 \cdot d + .5 \cdot .5 \cdot d$, see eq. (12)).

Right: (Solid line) ROC-curve for the conditional detection model (f_s^*, f_n^*). (Broken line) ROC-curve for the unconditional detection model (f_s^o, f_n^o).

A conditional high threshold model cannot be distinguished even theoretically from an unconditional high threshold model with smaller discriminability parameter. Consider e.g. the following functions:

$$f_s = d,\ f_n = f_g = 0, \qquad \text{for } b \leq 0,$$

$$\left.\begin{array}{l} f_s = d+(1-d) \cdot b, \\ f_n = b, \\ f_g = b, \end{array}\right\} \text{ for } 0 < b < 1,$$

$$f_s = f_n = f_g = 1, \qquad \text{for } b \geq 1.$$

The respective conditional model with m = 1

$$\begin{aligned} p(Y|s) &= f_s^*(d,b,p(o|s)) = p(o|s)\cdot(d+(1-d)\cdot b)+(1-p(o|s))\cdot b \\ p(Y|n) &= f_n^*(d,b,p(o|n)) = p(o|n)\cdot b+(1-p(o|n))\cdot b = b, \end{aligned} \qquad 0 < b < 1,$$

is identical with the unconditional model

$$\begin{aligned} p(Y|s) &= f_s^0(p(o|s)\cdot d,b) = p(o|s)\cdot d+(1-p(o|s)\cdot d)\cdot b, \\ p(Y|n) &= f_n^0(p(o|s)\cdot d,b) = b \end{aligned} \qquad 0 < b < 1,$$

with discriminability parameter $p(o|s)\cdot d \leq d$.

Obviously the relative frequency of guessing can be ascertained only if f_g differs clearly from f_s and f_n. Even so the contributions of perceptual decisions and guesses to hit and false alarm rates may not be easy to separate. Consider e.g. a model that assumes gaussian functions f_s, f_n and a step function f_g

$$\begin{aligned} f_s(d,b) &= {}^{b+d/2}_{-\infty}N(0,1), \\ f_n(d,b) &= {}^{b-d/2}_{-\infty}N(0,1), \\ f_g(b) &= \begin{cases} 1, & b > b_0, \\ 0, & b \leq b_0, \end{cases} \end{aligned}$$

i.e. unnoticed stimuli will be reported if and only if b exceeds some value b_0. In order to avoid the introduction of an additional free parameter, we may reasonably postulate that b_0 corresponds to pay-off conditions which yield a zero (or minimal satisfactory) expected gain for reports under a given a priori signal probability. The resulting functions of the conditional detection model

$$\begin{aligned} p(Y|s) &= f_s^*(d,b,b_0,p(o|s))=p(o|s)\cdot {}^{b+d/2}_{-\infty}N(0,1)+(1-p(o|s))\cdot \begin{cases} 1, & b > b_0, \\ 0, & b \leq b_0, \end{cases} \\ p(Y|n) &= f_n^*(d,b,b_0,p(o|n))=p(o|n)\cdot {}^{b-d/2}_{-\infty}N(0,1)+(1-p(o|n))\cdot \begin{cases} 1, & b > b_0, \\ 0, & b \leq b_0, \end{cases} \end{aligned} \qquad (13)$$

represent to the left and right of b_0, ogives which are reduced by multiplication with the respective inspection probabilities (see Figure 3). At b_0 both functions show a jump upwards of step size equal to the respective non observation probability. The ROC-curve of that model has three distinct parts. The upper and lower end are determined by f_s and f_n, whereas the middle part is a straight line of slope $1-p(o|s)/1-p(p|n)$. The difference between the $p(Y|s)$-values of the end points of the straight part equals $1-p(o|s)$, the difference between their $p(Y|n)$-values equals $1-p(o|n)$. (In Figure 3 two additional ROC-curves are shown for illustrative purposes. They differ from the main one with respect to the observation probabilities $p(o|s)$, $p(o|n)$ which were determined from the attention model $f'_s(c,a) = {}^{a+.22}_{-\infty}N(0,1)$, $f'_n(c,a) = {}^{a-.22}_{-\infty}N(0,1)$ by varying the attention parameter a.)

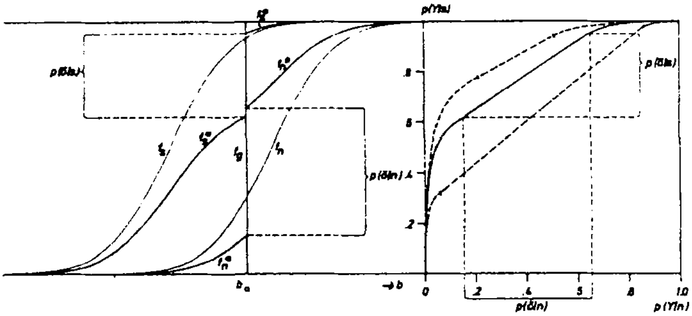


Figure 3

Conditional detection theory with Gaussian decision functions and a step function for guessing probability.

Left: Report probability for signal (f_s^*) and noise (f_n^*) as weighted mean of the Gaussian decision function f_s resp. f_n of Figure 1 and a step function f_g for guessing probability

$$f_s^*(b) = \begin{cases} .67\ f_s(b), & b \leqslant b_0 \\ .67\ f_s(b) + .33, & b > b_0 \end{cases}$$

$$f_n^*(b) = \begin{cases} .5\ f_n(b), & b \leqslant b_0 \\ .5\ f_n(b) + .5, & b > b_0 \end{cases}$$

Right: (Solid line) ROC-curve of the conditional detection model (f_s^*, f_n^*). (Broken lines) Examples of ROC-curves of the same model with higher and lower observation probabilities. (For values of $p(o|s)$ and $p(o|n)$, see Figure 4).

ROC-curves of conditional detection models with step functions f_g somewhat resemble the ROC-curves of unconditional two threshold models. If, however, a model with smooth functions f_s, f_n can be assumed - and models of that kind still have most empirical support (Laming (1973), p. 159ff) - then a model of conditional detection is identifiable in principle and may be identifiable in practice in spite of experimental error, provided a sufficient number of ROC-points, well distributed over the whole range of b, has been determined empirically. In a subsequent section we consider experimental conditions which favor the applicability of the model given in eqs. (13).

5. MODELS WITH CONSTANT GUESSING FUNCTIONS

Here, we consider models with constant probabilities of reporting unnoticed stimuli as signals. When assuming $p(Y|\bar{o}\wedge s) = p(Y|\bar{o}\wedge n) = p_g \equiv f_g(b) \equiv \text{const.}$, eqs. (7a) take the form

$$\begin{aligned} P(Y|s) &= f_s^*(d,b,p_g,p(o|s)) = p(o|s)\cdot f_s(d,b) + (1-p(o|s))\cdot p_g, \\ P(Y|n) &= f_n^*(d,b,p_g,p(o|n)) = p(o|n)\cdot f_n(d,b) + (1-p(o|n))\cdot p_g. \end{aligned} \qquad (14)$$

In view of the rather restricted plausibility of such a model with $0 < p_g < 1$, examples of the functions f_s^*, f_n^* and ROC-curves will be shown only for the special case $p_g = 0$ (see Figure 4). In the general case the f_s^*, f_n^*-ROC-curve is a reduced version of the f_s, f_n-curve which begins at the point $(1-p(o|n))\cdot p_g$, $(1-p(o|s))\cdot p_g)$ and ends at $(1-p(o|n))\cdot(1-p_g)$, $(1-p(o|s))\cdot(1-p_g)$. The values of $p(Y|s)$ are reduced by the factor $p(o|s)$, the values of $p(Y|n)$ by $p(o|n)$.

The most intuitive constant probability of reporting a signal on the basis of arbitrary guessing is

$$p_g = 0,$$

i.e. it is assumed that unnoticed stimuli will never be reported as signals. That assumption is most likely to be met by experimental data if signals are infrequent such as e.g. in vigilance tasks. In that case eqs (14) reduce to

$$\begin{aligned} p(Y|s) &= f_s^*(d,b,c,a) = p(o|s)\cdot f_s(d,b) = f_s'(c,a)\cdot f_s(d,b), \\ p(Y|n) &= f_n^*(d,b,c,a) = p(o|n)\cdot f_n(d,b) = f_n'(c,a)\cdot f_n(d,b), \end{aligned} \qquad (15)$$

the function f_s^* increases from zero to $p(o|s)$ and f_n^* from zero to $p(o|n)$, the ROC-curve from 0,0 to $p(o|n)$, $p(o|s)$. If signal and noise stimuli are equally conspicuous, i.e. if $c = 0$ and consequently $p(o|s) = p(o|n)$, the ROC-curve ends at the main diagonal.

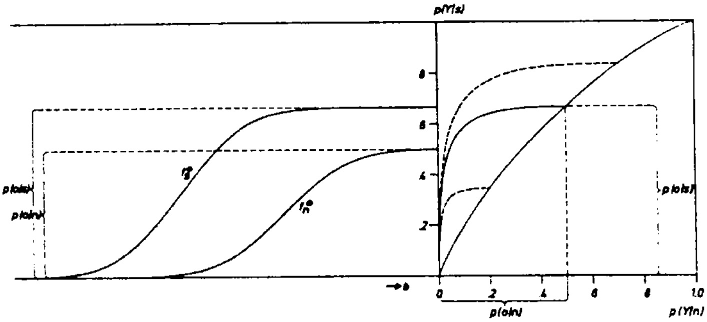


Figure 4

Conditional detection theory without guessing.

Left: Report probability for signal (f_s^*) resp. noise (f_n^*) as weighted mean of the Gaussian decision functions f_s resp. f_n of Figure 1 and the constant guessing function $f_g(b) \equiv 0$.
$f_s^*(b) = .67\ f_s(b)+0$, $f_n^*(b) = .5\ f_n(b)+0$.

Right: (Solid line) ROC-curve of the conditional detection model (f_s^*, f_n^*).
(Broken lines) Examples of ROC-curves of the same model with higher and lower observation probabilities.
The lowermost solid curve is the ROC-curve of a Gaussian attention model (f_s', f_n') with conspicuity parameter $c = .44$ from which the three pairs of observation probabilities $p(o|s)$, $p(o|n)$ have been computed ($p(o|s) = .34, .67, .83$, $p(o|n) = .2, .5, .7$).

In Figure 4, three ROC-curves are shown. They are derived from the Gaussian decision model with discrimination parameter $d = 2$:

$p(Y|o \wedge s) = f_s(d,b) = {}^{b+1}_{-\infty}N(0,1)$, $p(Y|o \wedge n) = f_n(d,b) = {}^{b-1}_{-\infty}N(0,1)$, combined with the Gaussian attention model with relative conspicuity parameter $c = 0.44$:

$$p(o|s) = f_s'(c,a) = {}^{a+.22}_{-\infty}N(0,1),\ p(o|n) = f_n'(c,a) = {}^{a-.22}_{-\infty}N(0,1).$$

They differ with respect to the attention parameter a. Their endpoints lie on the ROC-curve which corresponds to the attention model adopted.

Eqs. (15) have been fitted with some success to vigilance data (Tergan (1972);

Zimolong & Meyer (1974)). The model is identifiable, because, again, f_g is clearly different from f_s and f_n. That would be true also for $f_g(b) \equiv p_g = 1$, an analoguous assumption applicable to conditions of high a priori signal probability or very low cost of false alarms.

Eqs. (14) with $p_g = 0$ and $p_g = 1$ can be regarded as special cases of eqs. (13) with $b_0 = +\infty$, resp. $b_0 = -\infty$. In fact, that model turns out to be the only identifiable conditional detection model.

6. CONSEQUENCES FOR DETECTION AND ATTENTION PARADIGMS

The objective of detection theory is the separation and separate measurement of sensitivity and decision bias. If imperfect attention, too, is to be taken into account and to be measured, estimates of attention-, discrimination- and decision parameters can be obtained from a model of conditional detection, provided arbitrary guesses can be distinguished from perceptual decisions. In order to make them distinguishable,experimental and pay-off conditions should favor, since they cannot enforce, a particular response in the case of unobserved stimuli. At first sight an "undecided"-response associated with neither negative nor positive pay-off appears most natural. In fact, if the confidence rating method is employed and signal probability is .5, the utilization of a neutral category is indicated. That category will be used for responses based on decisions as well as guesses, but other categories are likely to be chosen only after inspection of the stimulus to be judged. The frequency of neutral choices attributable to perceptional decisions can be estimated by determining the best fitting ROC-curve of an unconditional detection model f_s, f_n, given the response frequencies in the non-neutral categories and an arbitrary portion of the frequency of neutral choice to be estimated as a free parameter. The remaining neutral choices can be attributed to perception independent guessing. Their relative frequency is an estimate of $p(\bar{o}|s)$ resp. $p(\bar{o}|n)$. If a priori signal frequency is larger or smaller than .5, the identification of a "guessing"-category will be much more difficult but may not be impossible altogether.

A neutral category cannot be used, however, in a Yes-No procedure because, in order to trace the whole course of an ROC-curve, strong positive (resp. negative) decision biasses must be induced. They require pay-off matrices that yield sufficiently positive expected pay-offs for reports, (resp. rejections) given an a priori signal probability. Under such pay-off conditions it is more profitable to report a signal (resp. to reject the stimulus) at venture rather than to give a neutral response or to suspense the decision. Thus, Yes- and No-responses based on perception independent guessing cannot be precluded by permitting 0 to avoid a positive or negative decision.

A way out of the dilemma is the separate empirical determination of the lower and upper part of the ROC-curve. For the lower part we may employ payoff matrices that prescribe gains for hits, losses for false alarms and zero pay-offs for misses and correct rejections. Decision bias may be manipulated by varying the gain/loss ratio, subject to the restriction that expected pay-off for reporting unobserved stimuli is (sufficiently) negative. Under that conditions O's can be expected not to report a signal on the basis of a pure guess. The ROC-points thus determined will rarely suffice for fitting the theoretical ROC-curve of an eqs. (15)-type of model. As a rule the estimates of the coordinates $p(o|s)$, $p(o|n)$ of the upper end of the ROC-curve will be unreliable too if only empirical points of the lower part of the curve are available for the fitting procedure. Therefore we have to determine the upper part of the ROC-curve by employing pay-off matrices that prescribe gains for correct rejections, losses for misses and zero pay-offs for hits and false alarms, subject to the restriction that expected pay-off for a stimulus rejection is negative under the given a priori signal probability. Under these pay-off conditions pure guesses can be expected to result in Yes-responses. A combination of the two procedures yields the uppermost and lowermost branch of a ROC-curve of the type shown in Figure 3.

By fitting a model of the type given in eqs. (13) with $b > b_0$ to the upper, and $b < b_0$ to the lower set of data points we obtain the wanted estimates of the attention and detection parameters. Here again, the decisive aspect of the procedure is the concentration of guessing responses on one response class, either reports or rejections. If guesses result in Yes- as well as No-responses, parameter identification will be virtually impossible.

REFERENCES

[1] Broadbent, D.E., Decision and stress, London, Academic Press (1971).

[2] Green, D.M., and Swets, J.A., Signal detection theory and psychophysics, New York, Wiley (1966).

[3] Jerison, H.J., Pickett, R.M., and Stenson, H.H., The elicited observing rate and decision processes in vigilance, Human Factors 7 (1965), 107-128.

[4] Laming, D., Mathematical Psychology, London, Academic Press (1973).

[5] Luce, R.D., Detection and recognition, in Luce, R.D., Bush, R.R., and Galanter, E., Handbook of mathematical psychology, vol. 1., New York, Wiley (1963).

[6] Swets, J.A., and Kristofferson, A.B., Attention, Annual Review of Psychology, 21 (1970), 339-366.

[7] Tergan, S.O., Vigilanz und Signalentdeckungstheorie (Vigilance and signal detection theory), Unpublished diploma thesis, Marburg University, Germany (1972).

[8] Wykoff, L.B., The role of observing responses in discrimination learning, in Luce, R.D., Bush, R.R., and Galanter, E., Readings in mathematical psychology, vol. 1, New York, Wiley (1963).

[9] Zimolong, B., and Meyer, F., Die Anwendung der Signalentdeckungstheorie auf Vigilanzdaten (The application of signal detection theory to vigilance data), Bericht über den 28. Kongreß der Deutschen Gesellschaft für Psychologie in Saarbrücken, 1972, Göttingen, Hogrefe (1974).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

IMPOSSIBLE OBJECTS AND INCONSISTENT INTERPRETATIONS

E. Térouanne
Université Paul Valéry
Montpellier, France

Since Penrose triangle and Escher drawings, a variety of "impossible objects" or "impossible figures" appeared, as well in graphic arts as in visual perception studies. Several attempts have been made to modelize such situations. The present paper proposes a new analysis in terms of a mathematical model of the consistency of an interpretation. The premise for it is that the impossibility, if any, is not to be searched in the drawing itself, but rather in the interpretation that an observer gives of it.

1. INTRODUCTION

The concept "impossible object" was introduced by L.S. and R. Penrose as "a special type of illusion", with reference to drawings of M.C. Escher. Ever since, numerous examples of "impossible objects" have been published, and several authors have made various attempts to analyse and modelize this concept (Cowan, Draper, Gregory, Huffman, Sugihara, Térouanne, among others). The common characteristic of these so-called "impossible objects" is to be two-dimensional drawings conveying the impression of quite familiar three-dimensional objects at a first glance, and giving the observer, possibly after a delay, a strong feeling of "impossibility" or "falseness". In fact, it looks difficult to provide any more formal definition of "impossible objects" or "impossible figures" (how could one define something which could not exist?...).

The first aim of the present paper is to claim that, if one intends to formalize such a situation for the sake of scientific analysis, it is better to replace the concepts "impossible something" by that of "inconsistent interpretation" of a figure.

What we mean here by "an interpretation of a figure" is any set of assertions about the elements of the figure, dealing with:

(i) the surfaces of the virtual object which would be represented by such and such an area in the figure, and their quality (planar, or convex, or concave,...),
(ii) the intersection lines or points between these surfaces which would be represented by such and such a line or point in the figure, and their concavity or convexity,
(iii) the occlusions, i.e. the lines in the figure along which one surface of the virtual object would hide another one,
(iv) and possibly additional properties such as parallelism between some faces, or collinearity between edges, and so on ...

In these terms, the situation usually described as "impossible object" is that of an observer who, having stated his own interpretation of a figure, detects a contradiction between some of his assertions or some of their consequences. Then we shall speak of "inconsistent interpretation" of the figure.

Thus the premise for the present work is that the "impossibility", if any, is something perceived by the subject, and not an intrinsic property of the figure. The following will examplify this main point.

First, it is well known, for instance since Necker's cube (fig. 1-a), that two observers of a figure, or the same observer at two different times, can give different interpretations of it. Another example is fig. 1-b which can be interpreted either as a hollow cube or as a square-based truncated pyramid built on the top of a cube (upper views).

Secondly, let us consider a variant of Penrose's triangle (fig. 1-c). The impression of impossibility occurs when interpreting it as the image of an object made of three parallelepipedic bars joining by pairs (that is what we call a "multibar object"). Nevertheless, alternative interpretations exist which are realizable as three-dimensional objects (Gregory, Térouanne). Moreover, some subjects do not feel any impression of impossibility with this figure, since they spontaneously interpret it as an image of a puzzle, or of an assembly of paper strips, and they can actually build such objects.

Finally, a geometrical argument (developed below) allows us to prove that fig. 1-d is actually the image of a multibar object, when most observers find

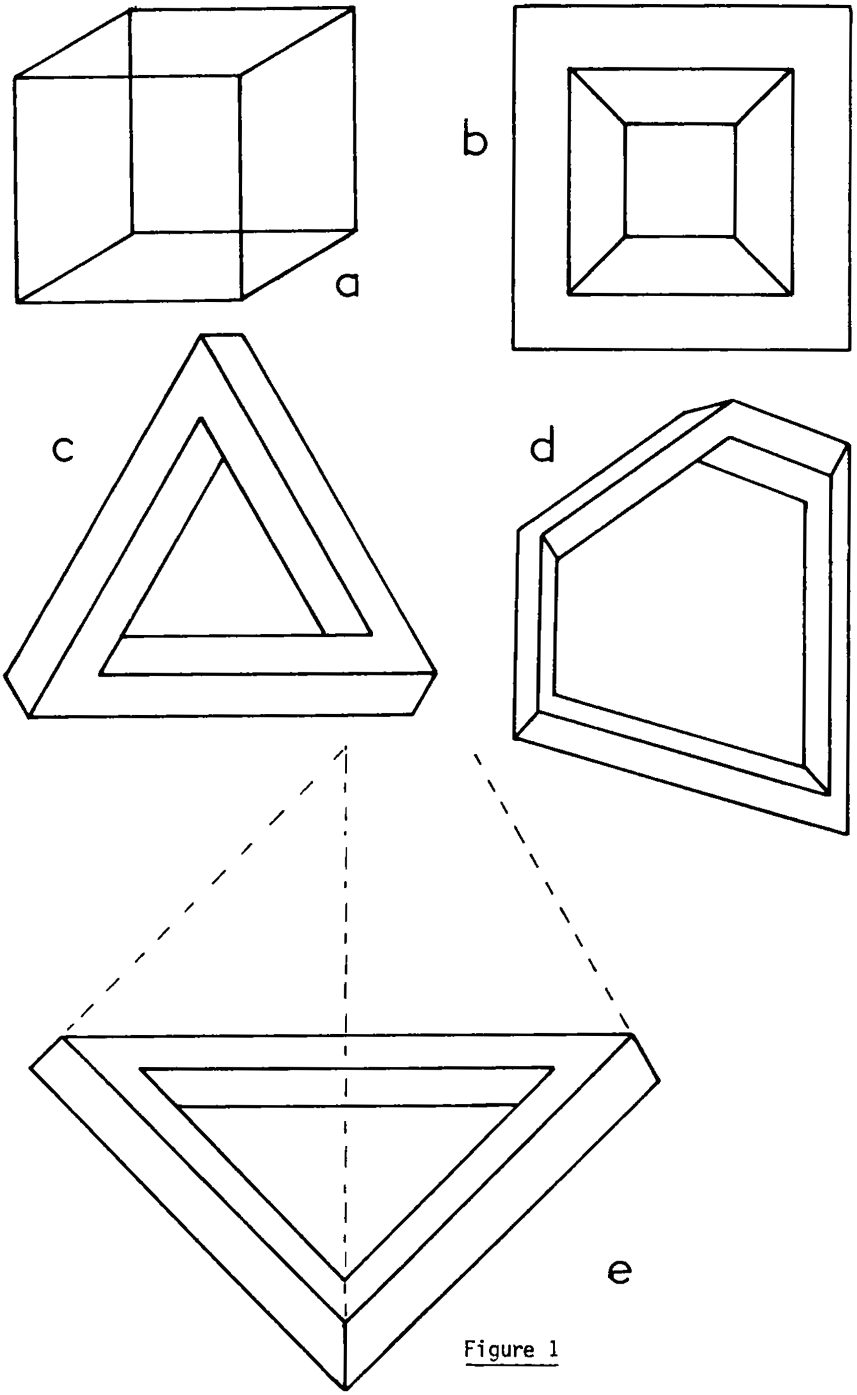


Figure 1

something "impossible", or at least "odd" in it. On the other hand, a simple geometrical argument proves that fig. 1-e cannot be the image of a multibar object (its three external folds should meet), when most observers interpret it as such. Here again, an alternative, and geometrically consistent, interpretation of the same figure 1-e can be given, for instance, by assuming that the left corner of the virtual tribar has been truncated.

We will now develop a mathematical model for the interpretations of a "polyhedral figure". When speaking of a "polyhedral figure", we assume that the figure is exclusively made of straight lines and that the observer interprets it as the image of a polyhedral scene, that is a scene exclusively made of objects with planar faces. Although it does not take into account all the situations usually described as "impossible objects", the family of polyhedral figures contains most of the known examples of them.

2. POLYHEDRAL SCENE

Consider a polyhedral scene and an observer of it: for instance in fig. 2-a the scene is composed of a triangular pyramid (with faces A,B,C) built on the ground D. Each face of the scene is supported by a plane of the space. Let F be the set of these planes (in our example F = {A,B,C,D}). Each visual ray of the observer is represented by an oriented half-line issuing from a given point O and going through the scene, much like X-rays. Then with each visual ray R we can associate a binary relation upon the set F, expressing the order in which the elements of F are met by R. For instance in our example, R meets A and B together at a point of their intersection, then it crosses the pyramid and meets the third side C before the ground D: we say that $A=B > C > D$ along R. On the other hand, the ray S crosses the same planes in the order: $A > D > B > C$.

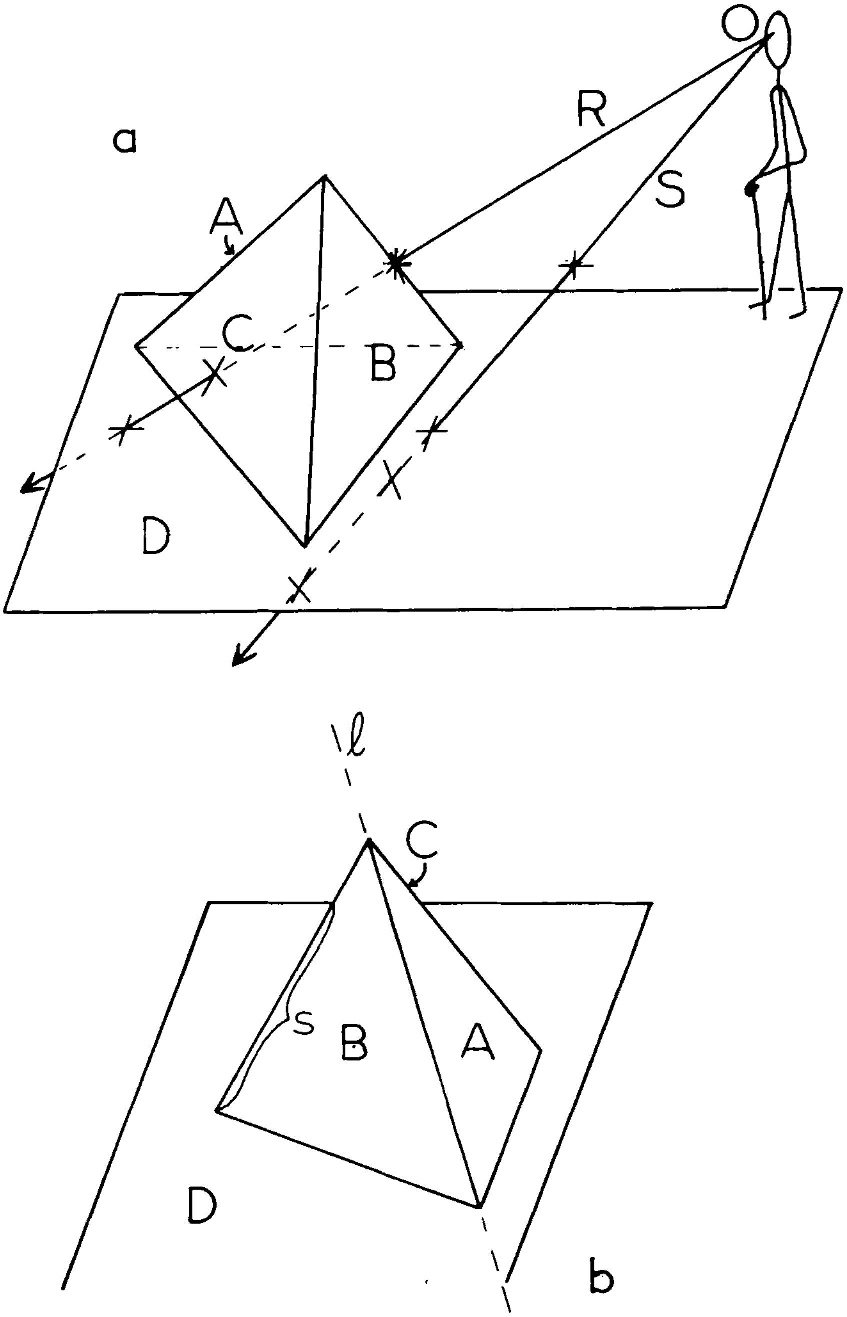

a
O
R
S
A
C
B
D
ℓ
C
s
B
A
D
b

Figure 2

3. POLYHEDRAL FIGURE

Let us now go back to a polyhedral figure and an observer of it. The observer is no longer represented in the figure since he is out of its support and is, typically, the present reader for fig. 2-b. Each point in the support of the figure represents one of the observer's visual rays towards the figure, and hence one of his visual rays across the space of the virtual scene. Whenever the observer interprets a line of the figure as the intersection of two faces, he splits the support of the figure into three parts according to the order in which the corresponding rays would cut the planes supporting these faces. For instance, if you interpret the line ℓ in fig. 2-b as the support of the convex intersection of faces A and B, you mean that at every point of ℓ, $A = B$ holds, whence at any point on the right side of ℓ, $A < B$ holds and at any point on the left side of ℓ, $A > B$ holds. Similarly, whenever the observer interprets a segment, say s, as an occlusion of a face, say B, over another face, D in our example, he assumes that at every point of s, $B > D$ holds.

4. INTERPRETATION OF A POLYHEDRAL FIGURE

Let S be the set of segments, half-lines or lines which compose a given polyhedral figure, and P the set of polygonal areas they determine in the support of the figure.

DEFINITION 1

An interpretation of a polyhedral figure (S,L) consists of:

(i) the set $\mathbb{F}$ of "faces" in the interpretation. Each face is related to a part of P. If this part is empty, the face is a "hidden" face.

(ii) the set $\mathbb{E}$ of "edges" in the interpretation. Each edge is related to one or several colinear segments or lines. Each of these segments or lines either belongs to S or is added by the observer with dots and is then called a "hidden edge".

(iii) the set $\mathbb{A}\subset\mathbb{F}\times\mathbb{E}\times\mathbb{F}$ of "adjacencies" in the interpretation. $(A,e,B) \in \mathbb{A}$ means that e is interpreted as a common edge between A and B and that the observer has decided (for instance by considering convexity or concavity) on which side of e, $A > B$ is valid.

(iv) the set $\mathbb{O}\subset\mathbb{F}\times S\times\mathbb{F}$ of "occlusions" in the interpretation. $(A,s,B) \in \mathbb{O}$ means that s belongs to an edge of A along which A occludes B.

(v) a set of equivalence relations on $\mathbb{E} \cup \mathbb{F}$ translating possible assumptions of parallelism, colinearity or coplanarity.

DEFINITION 2

The "adjacency graph" of an interpretation is the graph in which vertices are elements of $\mathbb{F}$, the edges are elements of $\mathbb{E}$, with the edge e joining the vertices A and B if and only if $(A,e,B) \in \mathbb{A}$.

DEFINITION 3

The "ordering graph" of an interpretation at a point M of the figure is the oriented graph obtained in the following way from the adjacency graph:

(i) if $(A,e,B) \in \mathbb{A}$ and M is on the line supporting e, then the edge e is removed from the graph and its vertices are merged.

(ii) if $(A,e,B) \in \mathbb{A}$ and M is not on the line supporting e, then the edge e is directed from A towards B (or B towards A) if the interpretation of (A,e,B) supposes that $A > B$ (or $B < A$) at M.

(iii) if $(A,s,B) \in \mathbb{O}$ and M is on s, then an oriented edge is added from A towards B.

5. LOCAL CONSISTENCY OF AN INTERPRETATION

DEFINITION 4

A point in the support of a polyhedral figure is a "critical point" for a given interpretation of this figure if and only if the ordering graph of the interpretation at this point contains a circuit.

DEFINITION 5

An interpretation of a polyhedral figure is "locally consistent" if and only if it allows no critical point.

For instance, fig. 3-b is the adjacency graph of an interpretation of fig. 3-a which involves only three visible faces A, B and C, and their paired convex adjacencies a, b and c. This interpretation proves to be locally consistent: as a matter of fact, the lines supporting a, b and c split the support of the figure into six areas, six half-lines and one point, and the ordering graphs in each of these thirteen regions contain no circuit: for instance, figure 3-c (or 3-d) is the ordering graph at point M (or N) but also at any other point of the same area delineated by the lines supporting

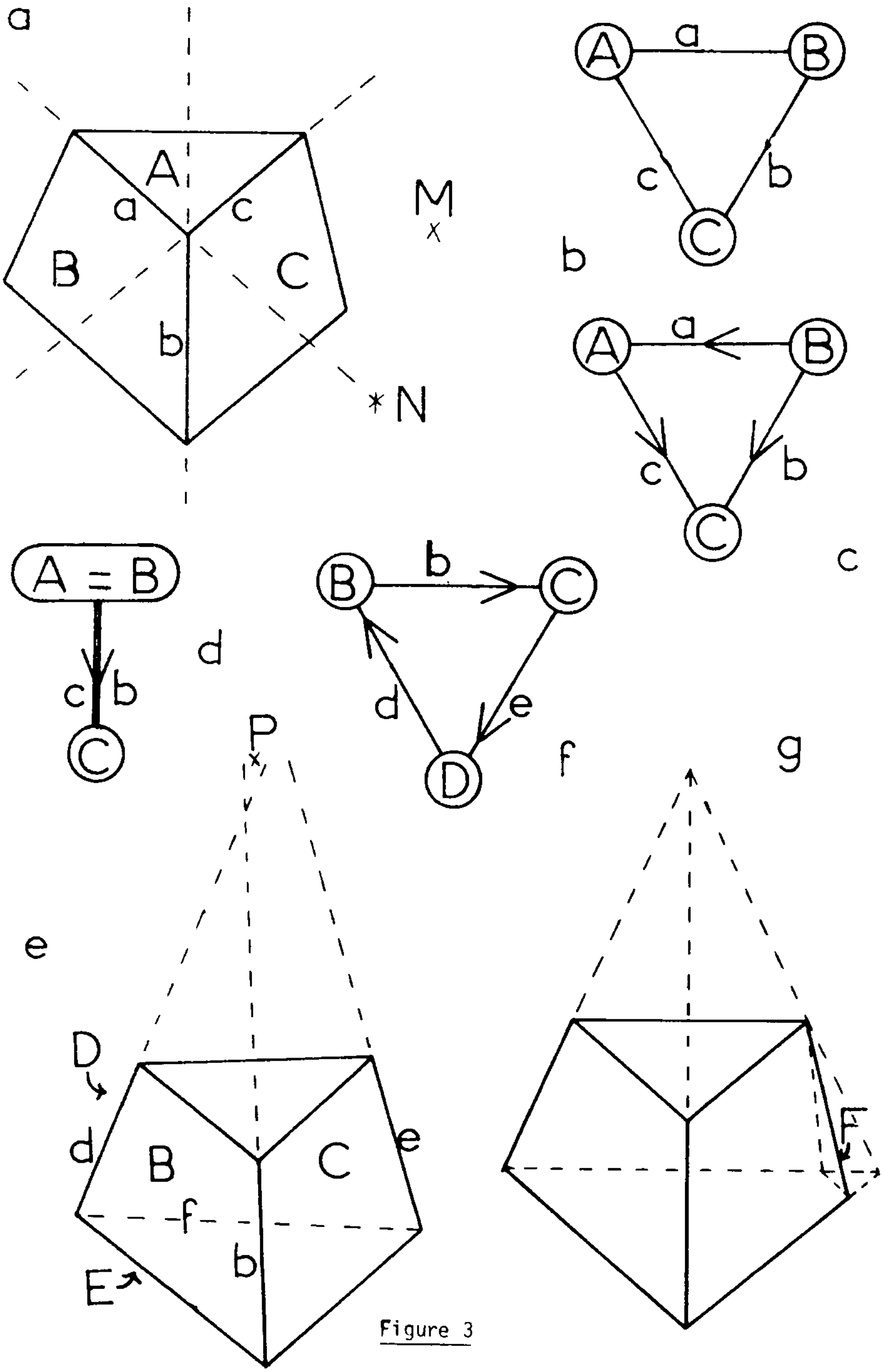

a
A
a
c
B
C
b
M
N
b
A
a
B
c
b
C
A
a
B
c
b
C
c
A = B
c
b
C
d
b
B
C
d
e
D
f
g
P
e
D
d
B
C
e
f
b
E
F

Figure 3

a, b and c.

The polyhedral figure in fig. 3-e is the same as in fig. 3-a, but we now consider its interpretation in terms of a truncated pyramid: two hidden faces D and E and a hidden edge between them appear. The adjacency graph now involves five faces and nine edges. The paired intersections b, d and e between faces B, C and D are not concurrent, thus the interpretation leads to a contradiction. As a matter of fact, the ordering graph at point P contains the circuit B>C>D>B (fig. 3-f) and P is a critical point for this interpretation.

Finally, fig. 3-g proposes a third interpretation of the same polyhedral figure, with an additional truncature of the virtual pyramid along a hidden face F. And this latter interpretation is locally consistent.

In order to check the local consistency of an interpretation, it is not necessary to compute the ordering graph at every point of the support, nor at every element of the dissection of the support by the lines supporting the edges of the interpretation. Consider for example the pentabar of fig. 4-a, and an interpretation of it restricted to its five visible faces and their paired intersections along the central edges of each of the five "bars". The adjacency graph contains but one cycle, which could provide at most two opposite circuits. The lines a to e have been oriented for the convenience of the demonstration: in order to make the circuit of fig. 4-b from the adjacency graph, we must situate ourselves in the half-plane to the left of a, because a is convex and we want A<B. Similarly, we must situate ourselves to the left of b, to the right of c, to the left of d and to the right of e. This intersection is empty. So is the intersection of the opposing half-planes. This interpretation is thus locally consistent. The same procedure applies to any interpretation after having listed all the possible circuits, and in particular all the cycles in the adjacency graph.

But the local consistency of an interpretation is not a sufficient condition for its feasibility: for instance, the Penrose "impossible staircase" of fig. 5-a, or the Huffman "intricated parallelepipeds" of fig. 5-b are locally consistent, when their "common sense" interpretations clearly contain contradictions and are infeasible. On the other hand, the "intricated cardboards" interpretation of fig. 5-c looks fair at a first glance, but the "triratio rule" introduced below will prove its infeasibility. In order to deal with this more evolved notion of consistency, we need some new tools.

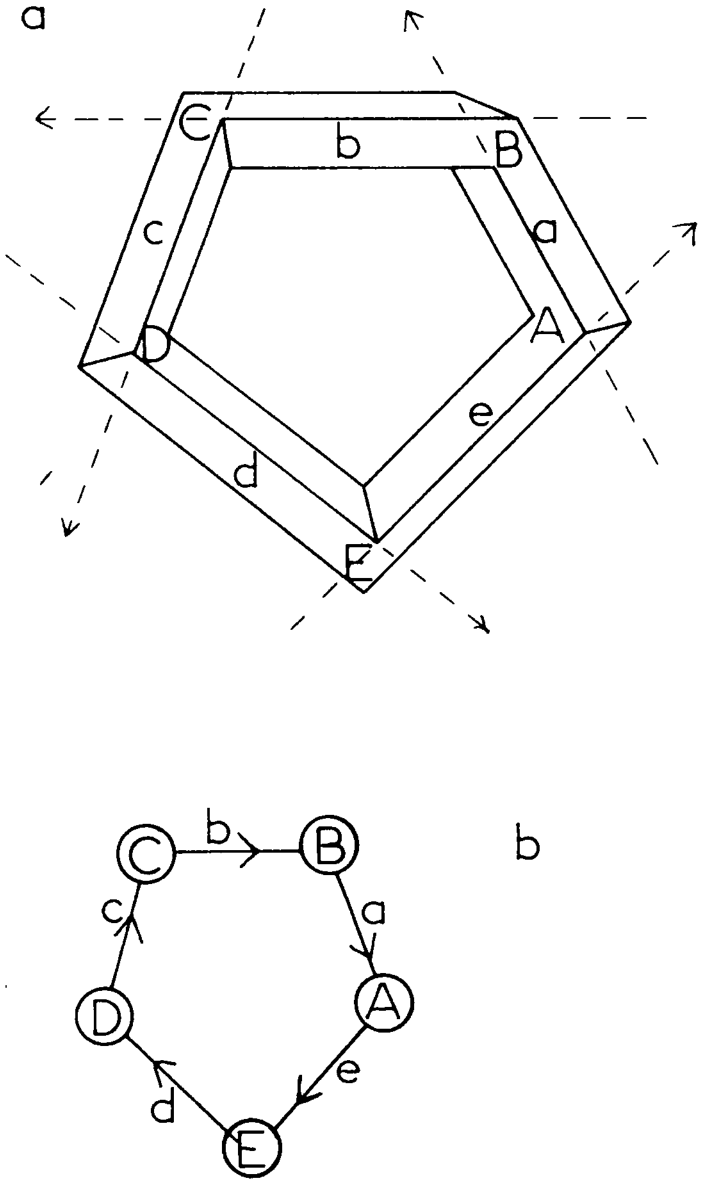

a
C
b
B
c
a
A
D
e
d
E
b
C
b
B
c
a
D
A
e
d
E

Figure 4

6. JOIN OPERATIONS

Consider the following operations defined in the Euclidean plane: if M and N are two arbitrary points in the plane, the "join" MN of M and N is the open segment between M and N, and the "extension" of M from N is the half-line M/N issuing from M and extending the segment MN. These operations are defined for the parts of the plane by union. They furnish it with a structure of "join geometry" or "join space", after Prenowitz.

Now, let A and B be two faces of an interpretation of a a polyhedral figure, and M and N be two points in the support of the figure. In order for the interpretation to be feasible, the orderings between A and B at M, at N and at some other points of the line determined by M and N should fulfil certain consistency relations: for example, if A>B holds at both M and N one should not find any point at which A<B holds between M and N. Or if A>B at M and A<B at N, one should find A>B at every point of M/N. And so on. All these relations are summarized in the following tables, where the empty squares correspond to the cases in which the orderings at M and N do not suffice to determine the ordering at the point considered.

N M	A>B	A=B	A<B
A>B	A>B	A>B	
A=B	A>B	A=B	A<B
A<B		A<B	A<B

N M	A>B	A=B	A<B
A>B		A>B	A>B
A=B	A<B	A=B	A>B
A<B	A<B	A<B	

Tables of the ordering between two faces A and B at an arbitrary point of MN (left) or M/N (right) as a function of their orderings at M and N.

7. TRANSITIVITY

Another feature of ordering relations which is clearly desirable in order to ensure the consistency of an interpretation is the transitivity: whenever A>B and B>C hold, one should also have A>C.

8. COMPREHENSIVE CONSISTENCY

Now we shall complete every local ordering relation by transitivity and by applying the consistency rules attached to the join and extension operations.

DEFINITION 6

The "comprehensive ordering relations" of an interpretation of a polyhedral figure at every point of its support are obtained from the local ordering relations at each point by alternately applying the two following procedures:

(i) transitivity: for every point M and every $A,B,C \in \mathbf{F}$ such that $A>B$ at M and $B\geq C$ at M, we add $A>C$ at M.

(ii) closure of the operations of join and extension: for any $A,B \in \mathbf{F}$, let us denote the set $\{A>B,\ A=B,\ A<B\}$ of the possible orderings between A and B at a point as $O(A,B)$. Then for every pair M, N of points in the support of the figure, in which the intersections of $O(A,B)$ with the current ordering graphs at M and N are non empty, we determine, according to the tables, which element of $O(A,B)$, if any, should be added to the ordering relations at every point of MN, M/N and N/M.

Because all the sets and relations in use here are finite ones, we can prove that the above procedure necessarily stabilizes at the end of a finite number (presumably small) of repetitions.

DEFINITION 7

The notions of "comprehensive critical point" or "comprehensive consistency" of an interpretation follow immediately. We shall eventually use "critical point" and "consistency", for short, and simply write "$A>B$ at M" for "$A>B$ belongs to the comprehensive ordering relation at M".

Here is for instance the demonstration of the inconsistency of the interpretation of fig. 5-b in terms of the imbrication of two arched parallelepipeds: (A,a,B) is interpreted as a convex adjacency and (B,KL,C) as an occlusion. It follows that $A>B>C$ at M. Similarly $A>E>C$ at P. Hence, $A>C$ at every point of MP. But $N \in MP$ and $C>D>A$ at N. Thus $A>C>A$ at N and N is a critical point.

The "staircase" interpretation of fig. 5-a assumes that all the "horizontal" parts of the "stairs", like X, Y and Z, are parallel. But whenever two parallel faces verify $X>Y$ at a given point, the join rules imply that $X>Y$ holds at every point of the support of the figure (for instance because we assume that $X=Y$ holds at some infinitely remote points). The demonstration of the inconsistency follows immediately.

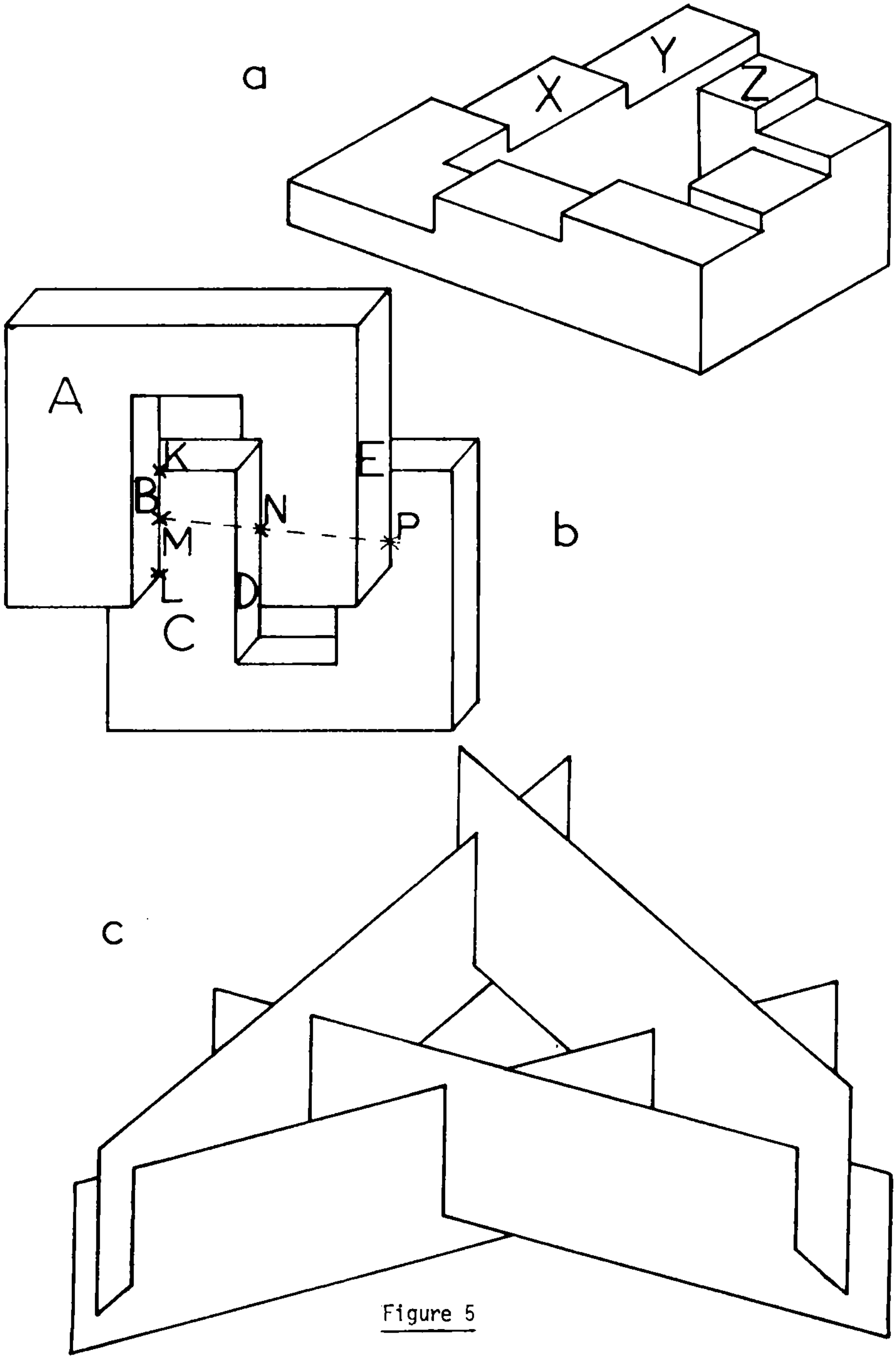

a
X
Y
Z
A
K
B
M
N
P
E
L
D
C
b
c

Figure 5

On the other hand, when an interpretation of a polyhedral figure is consistent, it often allows the observer to construct the point or line intersections of some triples or pairs of planes even though they had no common point in the figure itself. For instance, a dotted line in fig. 3-e shows the intersection line between faces C and D. And point P represents the intersection between B, C and D.

DEFINITION 8

An interpretation is "completable" when it is consistent and the comprehensive ordering procedure permits the construction of the line (or point) intersection of any pair (or triple) of its faces.

9. PARALLEL OR CONCURRENT INTERSECTIONS

The completability of an interpretation is not a sufficient condition for its feasibility in the special case, illustrated in fig. 5-c, where all the paired intersections between its faces are concurrent or parallel. Then we will need an additional property, related to the notion of "triratio" introduced by the following result.

THEOREM 1

If (D_i) $(1 \leq i \leq 6)$ is a family of concurrent or parallel lines in a plane, the quantity

$$[D_1,D_2,D_3,D_4,D_5,D_6] = \frac{\overline{A_1A_2}}{\overline{A_2A_3}} \cdot \frac{\overline{A_3A_4}}{\overline{A_4A_5}} \cdot \frac{\overline{A_5A_6}}{\overline{A_6A_1}}$$

is the same for every family (A_i) of colinear points in which $A_i \in D_i$ $(1 \leq i \leq 6)$.

DEFINITION 9

For any family (D_i) $(1 \leq i \leq 6)$ of parallel or concurrent lines in a plane, we shall call the "triratio" of the (D_i), the quantity $[D_1,D_2,D_3,D_4,D_5,D_6]$ defined by theorem 1.

THEOREM 2

Let (D_i) $(1 \leq i \leq 6)$ be a family of parallel or concurrent lines in the plane. Then the D_i are the projections of the paired intersections $F_0 \cap F_1$, $F_0 \cap F_3$, $F_0 \cap F_2$, $F_2 \cap F_3$, $F_2 \cap F_1$, $F_3 \cap F_1$ of four planes F_0, F_1, F_2, F_3 if and only if their triratio is -1. In this case we say that the lines (D_i) fulfil the "triratio rule".

10. FEASIBILITY OF AN INTERPRETATION

DEFINITION 10

An interpretation of a polyhedral figure is "feasible" when there exist a polyhedral scene and a bijective correspondence between the faces of the interpretation and a set of faces of the scene, and a bijective correspondence between the rays in the scene space and the points in the figure support, so that the ordering of any two faces of the interpretation at an arbitrary point coincides with that of the corresponding faces of the scene along the corresponding ray.

Now we can state the following result.

THEOREM 3

A sufficient condition for an interpretation to be feasible is that it is completable and, if all the paired intersections of its faces are concurrent or parallel, that they fulfil the triratio rule.

A more complete account of the present matter, with the proofs of all the results, is now available as a preprint paper.

REFERENCES

[1] Cowan, T.M., The theory of braids and the analysis of impossible figures, Journal of Mathematical Psychology, 6 (1977), 41-56.

[2] Draper, S.W., The Penrose triangle and a family of related figures, Perception, 7 (1978), 283-296.

[3] Escher, M.C., The world of M.C. Escher, New York, Abrams (1971).

[4] Gregory, R.L., The confounded eye, In R.L. Gregory, and E.H. Gombrich (Eds.), Illusion in nature and art, London, Duckworth (1973).

[5] Huffman, D.A., Impossible objects as nonsense sentences, In B. Metzler and D. Michie (Eds.), Machine intelligence, New York, Halsted, 6 (1971), 295-323.

[6] Necker, L.A., Observations on some remarkable optical phenomena seen in Switzerland, and on an optical phenomenon which occurs on viewing a figure of a crystal or geometric solid, London, Edinborough Philosophical Magazine, 1 (1832), 329-337.

[7] Penrose, L.S., and Penrose, R., Impossible object: a special type of illusions, British Journal of Psychology, 49 (1958), 31-33.

[8] Prenowitz, W., and Jantosciak, J., Join geometries: a theory of convex sets and linear geometry, New York, Springer Verlag (1979).

[9] Sugihara, K. , Classification of impossible objects, Perception, 11 (1982), 65-74.

[10] Térouanne, E., On a class of "impossible" figures: a new language for a new analysis, Journal of Mathematical Psychology, 22 (1980), 24-47.

[11] Térouanne, E., "Impossible figures" and interpretations of polyhedral figures, Multigraphié U.E.R. Mathématiques, Université Paul Valéry, Montpellier (1983).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

IN SEARCHING OF GENERAL REGULARITIES OF ADAPTATION DYNAMICS ON THE TRANSFORMATION LEARNING THEORY

V.F. Venda
USSR Academy of Science, Institute of Psychology

The systematic study of characterization equations, curves of strategies and of evolutions and transformations of strategies during the learning task are the fundamentals of a new transformation theory of learning processes. The theory has a theoretical and practical significance, particularly, for the solving of the problem of the optimal control of the learning process, and for more perfect experimental and analytical models of these processes. The individualization of the methods and means of the operator's learning, which becomes wide-spread nowadays, makes quite actual the development of a "micro-approach" to the study of learning processes, together with the traditional "macro-approach".

1. THE HISTORY OF THE DEVELOPMENT OF THE LEARNING THEORY AND THE CRITICAL ANALYSIS OF ITS PRESENT STATE

Traditionally, the problem of learning is one of the major problems of psychological science. Studies in this field were commenced by H. Ebbinghaus in 1885; he used the material of forgetting and remembering words. A great number of experiments gave him a possibility to find a monotonous exponential regularity for the decreasing of number of words one can store in his memory [17] .

Then, well-known E.L. Thorndike's experiments in the sphere of associations and Pavlov's studies aimed at the forming and destroying so-called "conditioned reflexes" in animals, gave the analogous results.

The greatest role in development of monotonous asymptotic models of learning has been played by C.L. Hull [[illegible]]. He assumed that the highest (asymptotic) level of skill is determined by intensity of need confirmation, conditioned stimulus and also by the temporal characteristics of stimuli and reactions. He invented the postulates about the generalizations of stimulus and it's dying away on the base of these postulates, and also of particular theorems on the learning to differ to structure stimuli, to reconstruct it volitionally, and so on C.L. Hull has gone to the main formula to calculate "the power of skill".

Later, after W.K. Estes', R.C. Atkinson's, G.H. Bower's, J.G. Greeno's, R.D. Luce's research, the monotonous asymptotic models of learning processes became classical and well-known. The first steps in the branch of learning psychology were fundamentally reviewed by G.W. Kling in Woodworth's "Experimental Psychology" [15], [20].

Significantly more than a half of quantitative experimental data about the learning processes were taken either from animals or from people, but using extremely simple tasks. Experimental situations with alternative reactions between 2 to 3 signals were most typical.

The large factual material about the learning processes gave a possibility for R. Bush and F. Mosteller [4] to build comparatively simple stochastic models of learning, which were based on the assumption that it's possible to describe these processes as Markovian ones.

Exponential models of learning processes found their way to applied psychology. They appeared to be useful for calculating the longitude of experimental series, in which the subject achieves effective solving for each from the finite set of problems; meanwhile this effectivity is comparable with the level of skilled specialists (but if it is a practically infinite set of possible real situations). This question is extremely important for the problem of transfering laboratory data onto real conditions of work.

The way of programming of engineering psychological experiments founded on using learning models which were proposed by W.K. Estes, R.C. Atkinson, R. Bush, F. Mosteller is described in our books [25, 31]. Shortly, this way reduces to the following things: beginning assumptions during the design and working as follows: 1) operative problems chosen are solved with the help of one certain method, i.e. all the ways and solutions are related with an unique strategy; 2) the learning process within the frontiers of this strategy is described as a Markovian one. J. Bruner talks about the strategy of decision making as on the certain way of acquiring, storing and using information which serves for achieving certain aims (results) [3].

The solution of problems by a subject means that his real strategy S_R is identical to an a priori strategy S_a, and the last one must be assimilated by the subject $S_R = S_a$, else $S'_R \neq S_a$.

When looking at the sequence of experiments as at the elements of a Markov process, we can see the following probabilities of transitions:

$P(S_R=S_a|S'_R\neq S_a)$, $P(S_R=S_a|S_R=S_a)$, $P(S'_R\neq S_a|S'_R\neq S_a)$, $P(S'_R\neq S_a|S_R=S_a)$.

Meanwhile, it is necessary to assume that right and wrong solutions cannot exist simultaneously in one experiment, and the sum of its probabilities equals 1; in particular, for beginning probabilities (before learning): $P_0(S_R=S_a)+P_0(S'_R\neq S_a) = 1$.

For the complete defining of the learning process as a Markovian one it is necessary to find, from experiments, following meanings: $P_0(S_R=S_a)$, $P(S_R=S_a|S'_R\neq S_a)$ and $P(S'_R\neq S_a|S_R=S_a)$; the other meanings of transition probabilities are easy to find from these ones.

We will symbolize $P(S_R=S_a|S'_R\neq S_a)$ as α and $P(S_R=S_a|S_R=S_a)-P(S_R=S_a|S'_R\neq S_a)$ as β; then the probability of the fact that in the training experiment number n, the subject would use the right strategy $S_R=S_a$ can be expressed as: $P_n(S_R=S_a)= \beta^n P_0(S_R=S_a)+(1-\beta^n)P_\infty(S_R=S_a)$, where $P_\infty(S_R=S_a)$ equals

$$\frac{\alpha}{1-\beta} = \frac{P(S_R=S_a|S'_R\neq S_a)}{P(S'_R\neq S_a|S_R=S_a)+P(S_R=S_a|S'_R\neq S_a)} .$$

Thus, if the probability of a right solution of each problem from the given set by the experienced operators, who were invited to take part in the experiments, $P_0(S_p=S_a)=P^*$, then from the above formulas, we can find the number of experiment n, in which a given subject who has certain initial training $P_0(S_R=S_a)$ and ability to acquire and store skills of right problem solving, $P(S'_R \neq S_a | S_R=S_a)$ and $P(S_R=S_a | S'_R \neq S_a)$, would achieve the stable level of results (the probability of right decision which is normal for skilled operators: $P_n = P^*$).

In connection with Bush-Mosteller's learning model, in the case of $-1<\beta<1$ and $n \to \infty$, the value of $P_n(S_R=S_a)$ goes to the asymptotic line $P(S_R=S_a)$, which reflects some finite constant level, the so-called "learning plateau".

The further development of the learning processes model was done by R.C. Atkinson, G.H. Bower, and E.G. Crothers, in "Introduction to mathematical theory of learning".

An important achievement of R.C. Atkinson et al., was the construction of the model of wholist learning process: the learning level in n was expressed rather from dynamics of learning, which was characterized by the variative velocity, than from the state during the experiment n-1 and constant transition probabilities from one state to another, as it was in Bush's models. Sequential learning of two-kind reactions is presented by R.C. Atkinson as two exponential curves: $P(S'_R \neq S_a) = (1-c)^{n-1}$ and $P(S_R=S'_a) = (1-qc)^{n-1}$, where (1-c) is the logarithmical decrement of dying away of wrong solutions probability; (1-qc) is the less decrement of dying away ($q \leq 1$) of the probability of strategy acquired in beginning S_a, which would be substituted by strategy S''_a. The probability of the fact, that in experiment number n would be observed the strategy $S''_R=S''_a$ is given by $P(S''_R=S''_a)=1-P(S'_a \neq S_R)-P(S_R=S'_a)$.

In the beginning $P(S_R=S'_a)$ is not high, then, when we are learning, it increases, achieves a maximum and then begins to be substituted by the second kind of reactions - by the strategy $S''_R=S''_a$; in a meantime $P(S''_R=S''_a)$ increases in connection with the above mentioned formula.

Using the control experimental data R.C. Atkinson had shown the dependence of the relative quantity of dogs who avoided the electrical current shock in experiments conducted by Solomon-Winn. These fact were used by Atkinson to prove coincidence between his mathematical models and experimental results.

Although R.C. Atkinson and col. generalized in their mathematical models practically exclusively the data of simple experiments with animals, these models reflected extremely important aspects of learning - the transition process from one type of acquired reactions to another - and marked the development of non-monotonous mathematical learning models.

Non-coincidence between real learning processes and monotonous exponents was shown in the wide range of experiments, but didn't find enough theoretical proofs. On the base of the great number of experimental data H. Ebbinghaus and later R. Woodworth differed initial, intermediate and final plateaus in the learning curves.

An intermediate plateau was firstly found experimentally by Brien and Harter in 1896 when studying processes with telegraphists. And it was interesting that the learning curve for the simplest operations was monotonous and hadn't an intermediate plateau, which was beautifully seen in receiving cable messages learning curves [6] .

Brien and Harter explained the appearance of intermediate plateaus during the learning, by the presence of forming elementary skills needed for further development, more complex and perfect skills (fig.1). Also they registrated the fact that the reception differs from translation, because in the first case a telegraphist can not vary the temporal characteristics of his work, because they are given by the objective characteristics of the external world. A plateau is a comparatively long delay in increasing the professional results, connected with the acquiring of wide-spread verbal lore, and then the process of learning makes a big leap. Brien and Harter showed in 1899 that the plateau is still more clear when learning more complete skills, for instance, when learning a reception connected text (fig. 1, curve 3), meanwhile when learning a reception words it's expressed less clear, and when learning a reception independent letters there isn't any plateau in general (curve 1). In the last case the learning process has a monotonous character. In 1906 Swift has published the data about the presence of an intermediate plateau when studying foreign languages.

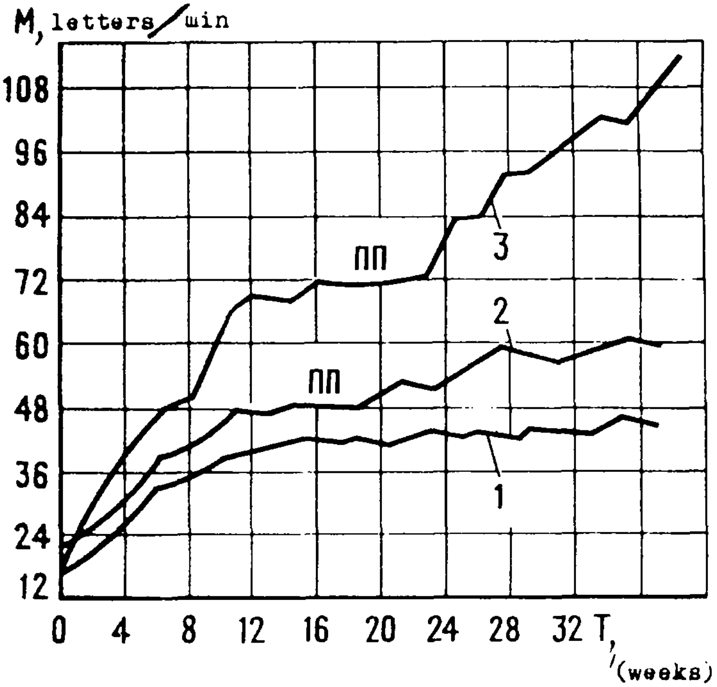


Figure 1

The curves of learning of telegraphists (by Brien and Harter, 1899).

Betson in 1916 and Kow in 1937 supported the conception that the formation of an intermediate plateau in learning curves is more probable with complex operations than with simple. Furthermore, they showed that complex operations, acquiring as a whole, did not give a plateau imperatively; they met it usually in those cases, when the student fixed his attention on the single elements of a complex operation, one after another.

If earlier monotonous learning processes appeared more often in animals and resultively had analogous processes when learning people to fulfil simple tasks, then later on it was made a successful trial to reproduct non-monotonous learning processes in animals by means of complicating their behaviour and learning conditions. For example, Hanter (1929) found a plateau when learning rats to search a way within the labyrinth.

Generalizing the results of all the experimental studies where intermediate plateaus were found, R. Woodworth explains this phenomenon by means of the changing of fulfilling task method. The analogous change was found by K.K. Platonov and L.M. Schwarz (1949), when they worked out motor skills, and the subject turned from surplus, bed regulated motions into more laconic purposeful motions. During the training two plateaus were observed: 15th to 18th and 26th to 28th days of shot's training. Among other causes of plateaus rise, they emphasize the influence of changing skill structure, turning into the skill with more perfect structure. For all that, it's possible a provisional change for the worse of the indexes which follows plateau.

B.F. Lomov [21] in his book has shown the convincing data about the non-monotonous character of learning processes and its dependence on the level of complexity of skills acquired (fig. 2).

The publication of such data was extremely interesting, because the generally accepted psychological and mathematical learning theories dictated almost obligatory an approximation of experimental data by means of a monotonous asymptotic learning curve. For instance, in the well-known handbook of A. Chapanis, devoted to methods and techniques of engineering-psychological experiments, namely this approximation of experimental data by an asymptotic learning curve was recommended. Fig. 3 taken from this book, shows an experimental point and "average curve" (unbroken line), proposed by Chapanis. It's possible, however, to put forward an alternative assumption as to the more complex character of a learning process (additional touch line at fig. 3). It's also interesting to note that in another chapter of the handbook, A. Chapanis presents non-approximated data and, however, doesn't give an explanation of the complex character of this process [16].

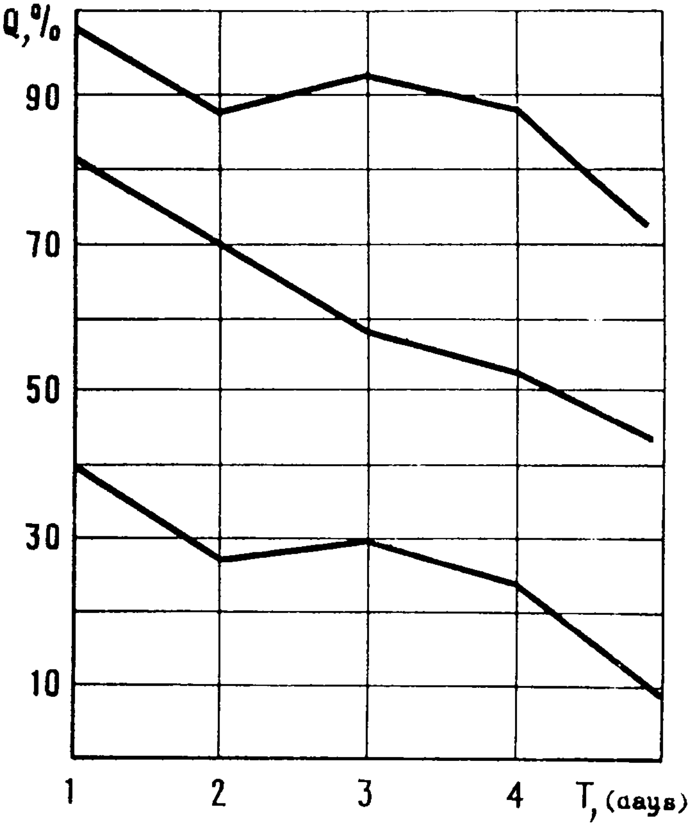


Figure 2

Examples of non-monotonous curves of learning of operators (by Lomov, 1966).

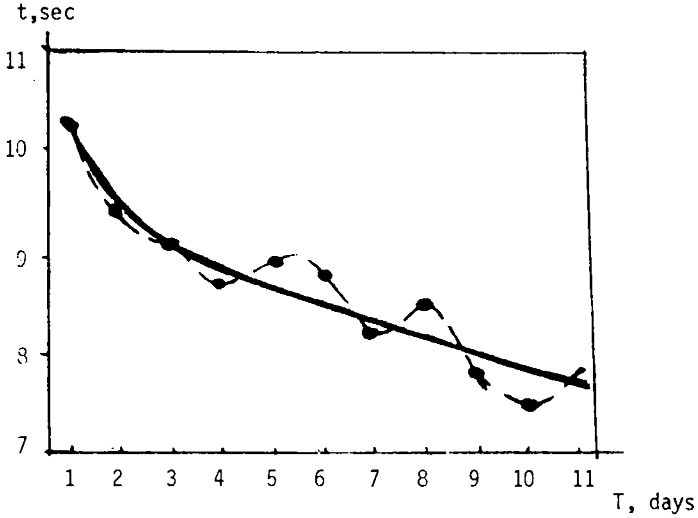
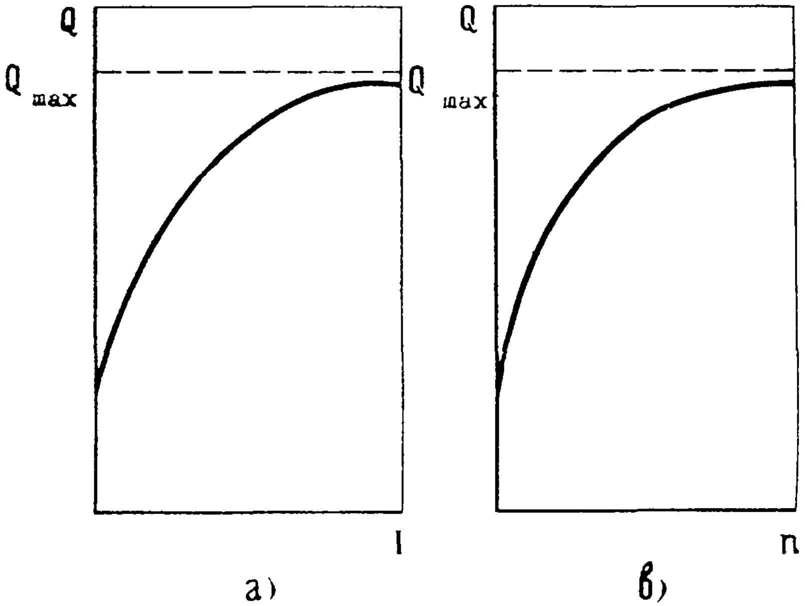


Figure 3

Experimental dots and their approximation by an exponential learning curve (by Chapanis, 1959). t : average keying time; T : days of test; . : experimental dots; — : Chapanis' approximation line; --- : additional approximation line suggested by us.

An approximation of the learning processes by the monotonous exponential curve became the most stable tradition in some applied branches of psychology and in ergonomics. For example, in the competent work "Introduction to ergonomics" (Moscow, 1977) mathematical methods of computation curves, proceeding from three or even from only two points are reviewed. Authors showed the monotonous dependence of the operator's job effectivity Q upon the volume of information I used and the corresponding monotonous learning curve asymptotically approaching to some "maximal" level of effectivity (Fig. 4), if the number of learning trials n is big enough.



a) б)

Figure 4

Effectiveness Q as a function of information volume I(a) and number of trials n(b). (by Zarakovsky et al., 1977).

Because analysis methods of learning processes presented above reflect a modern state of learning theory and are proposed by authors for the aims of wide practical application, let's discuss in detail these materials.

The authors proceed from the so-called "macro-approach", which is based on a generalized analysis of a great number of experimental data and their mean quantities. No doubt that such approach gives a possibility to show sometimes the most generalized regularities, but in this case it may lose the very significant traits of learning process. Let's illustrate this fact by the following example. Our experimental data (Venda, 1969) about the learning subject in compensatory tracking of dynamic signals are shown in Fig. 5 and an identical flow of watched signals, were output to different numbers of apparata (n=1÷6) which he could deal with simultaneously and send

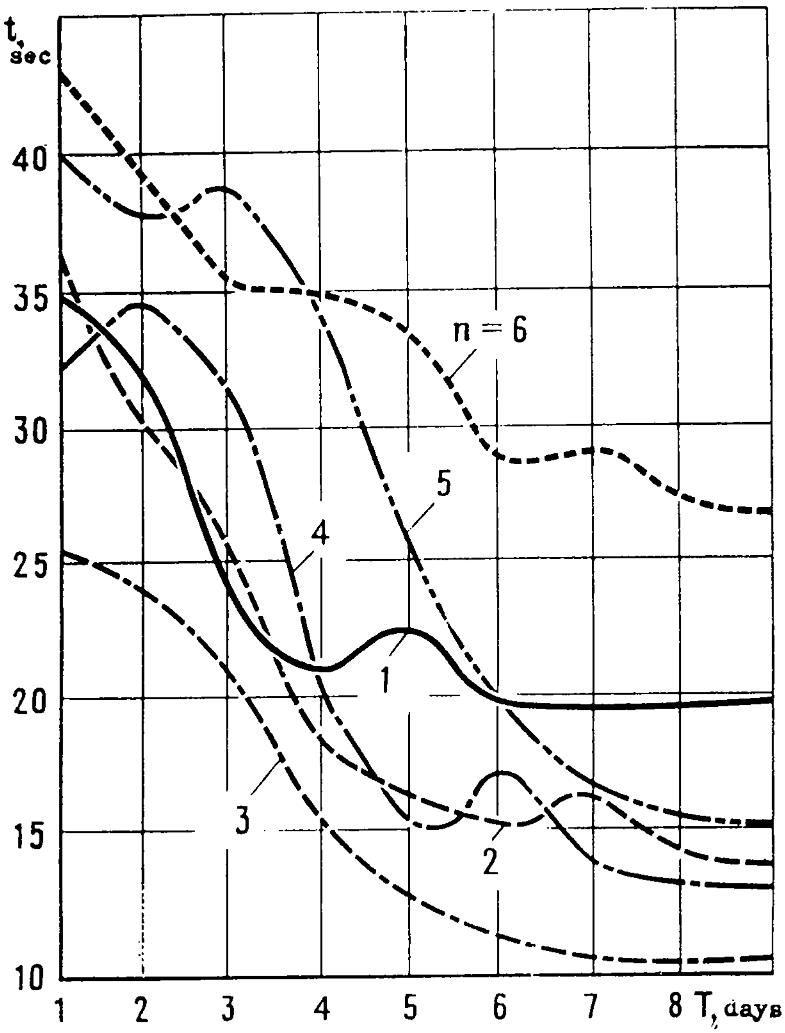


Figure 5

Experimental data on learning process with compensatory tracking : t : average time of tracking of one signal; T : days of learning; n : number of simultaniously perceived signals (n = 1; 2; 3; 4; 5; 6). (by Venda, 1969).

with the help of corresponding channels the regulative signals.

The process of learning appears to have openly a non-monotonous character when any number of apparata exists; it includes different platforms and even in some period shows the change of watching indexes to the worse. If all these complex curves are summarized arithmetically, it would be quite the traditional monotonous average curve (see Fig. 6).

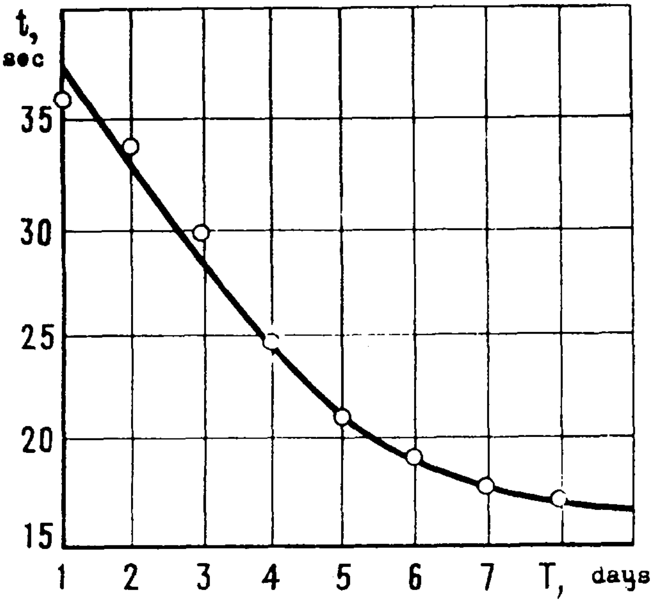


Figure 6

Total summarized and averaged data of Figure 5.

For all that, however, we lose the presentation about the differences during the learning with the different quantities of apparata, intermediate and final effectivity of watching, depending on the number of simultaneously receiving signals. After all, when using "macro-approach", we cannot analyse the influence of the informational factor on the learning, control the learning, solve the problem of adaptation of informational flow to operator by means of regulating a number of signals which are to be received by him simultaneously.

Yet at the beginning of our century, the scale of relative time with the registration of individual differences in learning speed and the corresponding stretching and compression of time axis during the joint reflection of data in the experiments with the different people or under different circumstances of their learning, aimed at approximation of curves, was proposed. The scale was named after its author, S.B. Vincent, who used successfully this scale when generalizing different data about the learning; however, then the scale applied very seldom.

If we would use S.B. Vincent's scale to find mean quantities of the same data, marking at the axis of relative time $T_{rel.}$, the co-ordinates of the different phases of the learning process, we would have a non-monotonous curve (Fig. 7), in which for each strategy, acquired in succession, it is possible to watch for the following phases : the beginning of strategy acquiring (1, 2), swift improvement of indexes (2, 3), plateau (3, 4), turning from the preceding strategy into the following (4, 5).

An analysis of such a curve gives us a possibility to ask a question, not only about isolated studies of the processes of human learning to each strategy as such (as it was done by Brien and Harter), but also about the studying of turning one strategy into another. It gives us a right to research not only the evolution of single strategy, but also the transformations of strategies.

Up to the time such transformations - as it seems to me, the most interesting - often were out of mind.

Authors of "Introduction to ergonomics" kindly presented to me their real data about some complex processes of learning which were not included in their book, because they were considered to be "artefacts" and "blunders". In one case, this changing of the number of misses admitted by operators when fractioned useful signals concerning the noises (see Fig. 8).
For all that, during the sixth hour of learning, it was proposed to the subjects to continue their work independently, without instructions. The student turned into the independent working out and defining the criteria

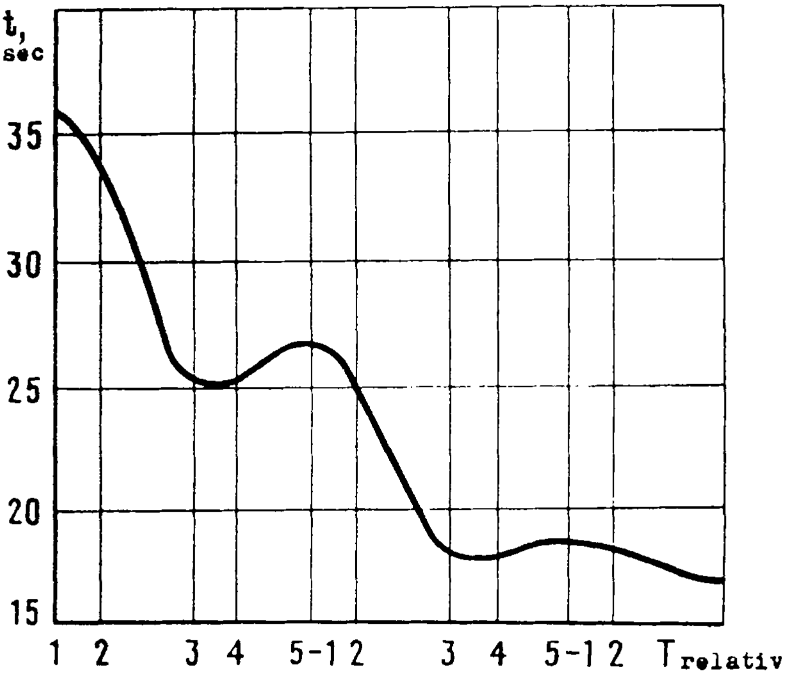


Figure 7

The same experimental data (of Fig. 5) summarized and averaged with scale by Vincent,(1912).

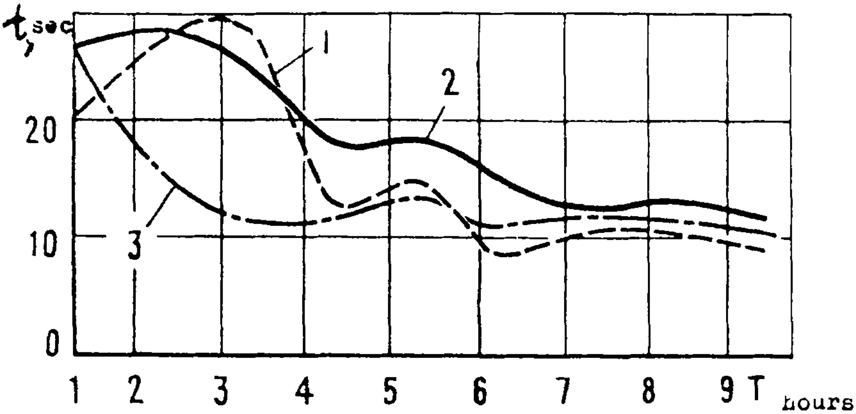


Figure 8

Real experimental data (compare with theoretical curves of the same authors at the Fig. 4). (by Zarakovsky et al., 1977).

of the fractioning useful signals. These quite numerous data cannot be approximated by monotonous curves for enough reasons.

In R. Atkinson's et al. book "Introduction in mathematical theory of learning" is presented the dependence of latent period of subjects responses t upon the quantity of preliminary remembered objects d; and for all that the segments for "small" and "big" number of objects are taken, but the intermediate numbers are eliminated from the consideration (see Fig. 9).

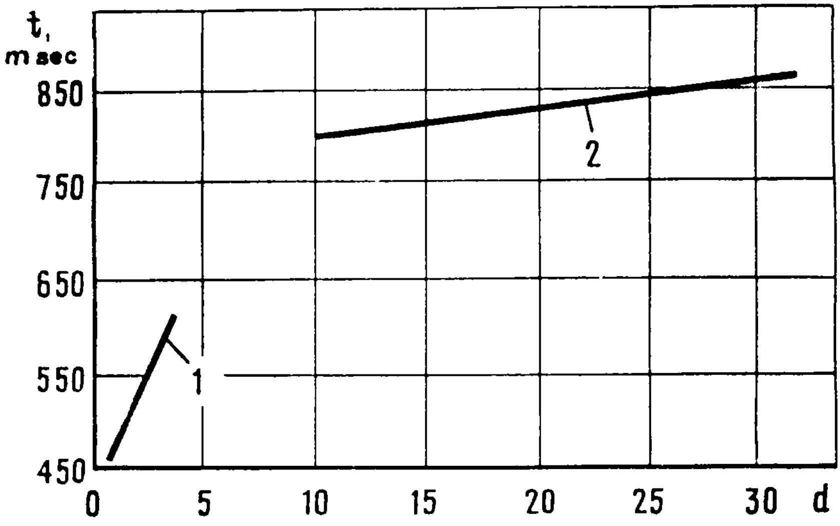


Figure 9

by Atkinson et al.(1971).

The dependence of t on d has a clear linear character. It's easy to find a monotonous asymptotic character of the learning, if we would assume regular monotonous relation between the characteristics of complexity and the indexes of successfulness, complexity of it's fulfilling in each moment of learning.

We must especially remark the question about interrelations between the factors (initial conditions) and indexes (results) of fulfilling tasks.

For instance, A.N. Leontyev and E.P. Krinchik (1963) showed a great set of data about linear relation between the informational capacity of signals (as psychological factor of task fulfilling complexity) and time of reaction to signals (as the index, criterion of task fulfilling complexity).

An analogous linear relation - the less information in signals, the easier the task fulfilling - cannot be observed, if we watch for real activity when achieving concrete aims.

For example, when solving some thinking or senso-motor problem, a human operator must have certain volume of information. When the volume has an optimal meaning, the problem is solved easily, but if the volume of information is more or less than optimal, then the complexity of solving would increase.

Fig. 10 shows an experimental curve of the dependence of the meantime of the compensatory watching for signals in the case, in which the signals composing the same flow are reflected in 1-6 apparata by turns (n=1-6). This curve is built in connection with the data of Venda, 1969. (see fig.10) Fig. 5 clearly shows an optimal number of apparata, which gives a possibility for meantime of the watching of signals to be minimal t(n=3)=11 sec., $n_{opt.}=3$. On the left ($n \leqslant 3$) and on the right ($n \geqslant 3$), the curve is approximated by the different regression equations : $t_1(n \leqslant 3)=9+(4,5-n)^2$ sec; $t_2(n \geqslant 3)=11+(n-2,5)^2$ sec.

The analogous curves with the clearly expressed optimum for many cases of connections between the different psychological factors and criteria of thinking, perceptive and senso-motor problems solving complexity are found. For all this, to each problem solving strategy, corresponds a certain kind of curve and certain optimal meaning of the psychological complexity factor. For simplicity, let's review a monofactor dependence; the methods of finding out of the system of psychological factors of complexity and their influence onto criteria of complexity chosen are discussed in detail in our monograph "Engineering psychology and synthesis of information display systems" (Moscow,(1975);2nd ed. -(1981)).

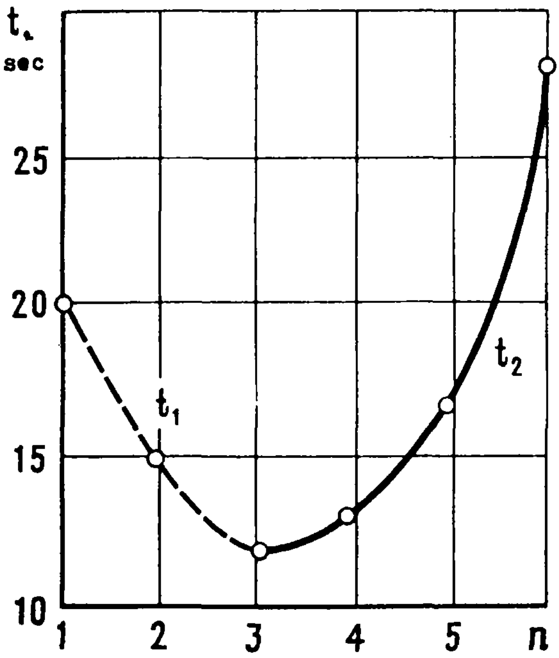


Figure 10

Characterization curve of strategy in compensatory tracking (by Venda, 1969). t : average tracking time; n : number of perceived signals.

Thus, there is a special aspect in the problem of learning to be discussed, and this aspect is closely connected with the emphasizing non-linear character of connection between the factors (conditions) and criteria of the successfulness of human task fulfilled during the learning and also of the presence of an optimal factor index for each task when fulfilling a real strategy in each experiment.

Although yet in 1896 the complex non-linear components in learning dynamics were found (and particularly, intermediate plateaus were observed) and the differences between possible strategies of fulfilling educational tasks were proved, there are practically only monotonous models of learning in contemporary psychological and mathematical learning theories.

Psychologists-mathematicians selected from the experimental-psychological body of facts about learning the most suitable ones for approximation by the simple analytical functions-Markovian chains, exponential and power equations, polynomials and other mathematical models, which are quite understandable and acquirable for all the members of research of the complex problem of learning, and particularly, for experimental psychologists. Such a material includes in main the processes of animal learning but also the processes of acquiring exclusively simple skills by humans.

Then, the experimental psychologists who studied more complex learning processes, often tended to represent the results of their researches by well-known analytical functions. For all this, the data which were not consonant with monotonous asymptotical and other analogous mathematical models, were eliminated as "artefacts".

The smoothing and "standardization" of learning curves were caused also by averaging of experimental data in connection with the great number of subjects and under the different circumstances of learning.

The monotonous character of learning curves was determined also by the assumption about the monotonous (usually, linear) character of dependence between the conditions (factors) and the criteria of task fulfilment complexity.

When studying the learning processes in engineering psychology and ergonomics, the scientists confined theirselves only with the "macro-approach" and didn't penetrate to the structure of processes, didn't find out the leading psychological factors. Last ones determine the regularities of learning and the characteristics of the learning processes which consequently make themselves familiar to destroy, change by more adequate ones and perspectives of the problem solving.

The further development of the learning theory is needed to be connected with the studying not only of the strategies evolution processes, but also of the processes of its transformation, turning from one strategy to another under the influence of learning methods, individual differences between the students, conditions and means of fulfilling educational tasks.

2. SOME WAYS OF FURTHER DEVELOPMENT OF THE LEARNING THEORY (INTRODUCTION TO THE TRANSFORMATION THEORY OF LEARNING)

An analysis of dependence between the psychological factor and the criterion of task fulfilling complexity, characteristic for each strategy, and also an examination of optimal index of this factor, gives us a possibility to define top possibilities of the perfection for each strategy and to determine the character of strategies transformations, i.e. of the turning from one into another.

Let's review an example of experimental data about the dependence between the psychological factor and the complexity criterion when fulfilling tasks and using the different strategies of its fulfilment.

Detailed descriptions of such experiments are given in [22, 25, 31]. Fig. 11 shows on the left the dependences of the operative problems solving time t upon the number of single informational signals n, received by the operator, when solving problems in mnemoscheme. Strategy A which begins the learning has a character of element-to-element perception of information from the mnemoscheme. The minimal solving time when using a strategy A t_{Amin}, we have in the case of $n^{A}_{opt.}$.

During the following stage of learning, the operator acquires a strategy B, in which the elements perceived are joined into wholist functional groups : t_{Bmin} in the case of $n^{B}_{opt.}$.

The operator achieves the best indexes when learning to perceive all the mnemoscheme simultaneously and to identify the state of object as a whole : t_{Cmin} $(n^{C}_{opt} = 1)$.

It's necessary to draw attention to the special points on the left in the curves of fig. 11, the common point AB of strategies A and B and the common point BC of the strategies B and C. Co-ordinates of these points correspond

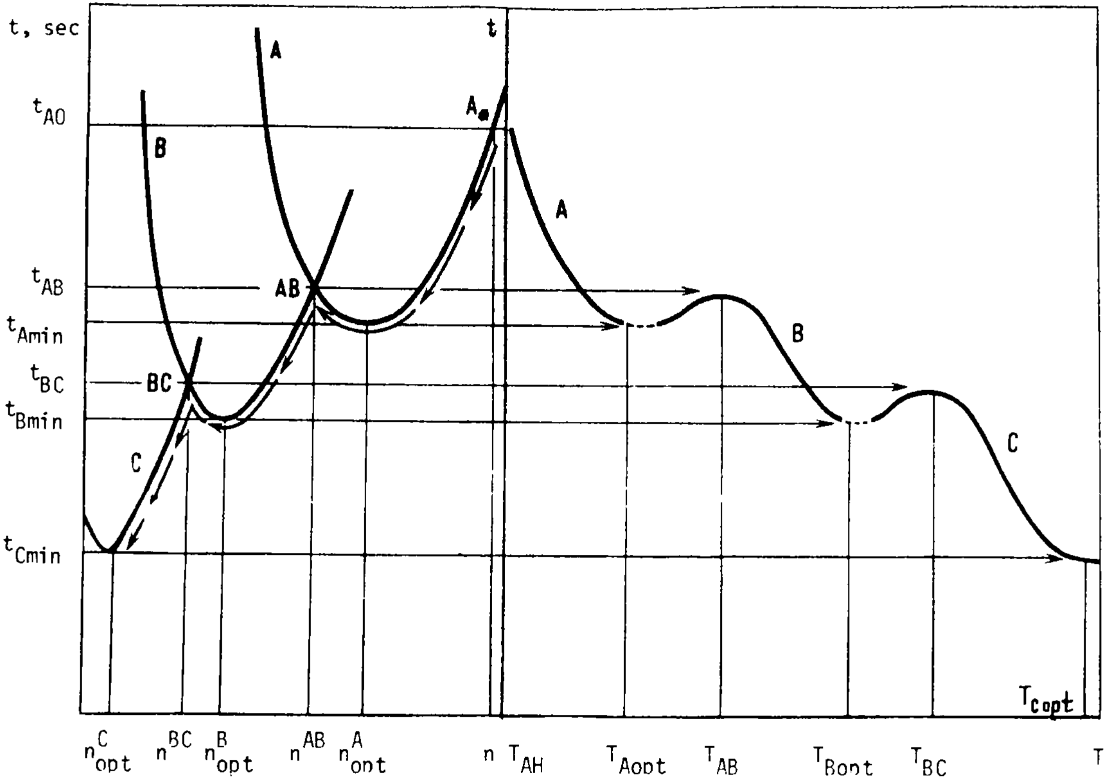


Figure 11

Characterization curves of strategies A, B, C (left) and learning curve (right) for criterion of time of decision making.

to identical indexes when using two different strategies. These points can be examined as the turning from one strategy into another, more perfect, when learning (from A into B and from B into C). This phenomenon can be partially explained on the base of the data about the transfer of skills.

Now the natural question arises : how the strategy-change gives us a possibility to decrease an optimal level of complexity $t^A_{min} > t^B_{min} > t^C_{min}$, maximally achievable when using each strategy.

It's necessary to make a reservation that in the given case we mean one of four levels of experimental data averaging : total (for all men), contingental (for certain mainland of students), group (for group of people from the students contingent who are united by some common characteristics) and then the individual level of the analysis of experimental data concerning the studying concrete man. We have there the main levels usually shown when we talk about the multi-level mutual adaptation of man and machine in the man-machine systems [21, 31]. The learning is between the main components of such an adaptation. For all that, when reviewing evolutionary and transformational processes in the field of learning strategies we also draw our attention to the mutual adaptation of man and learning conditions, means, instructions. Particularly, the selection of optimal, individually adapted to the student means of the fulfilling educational tasks, gives us a possibility to make the transformation from one strategy into another, simple and to acquire them till the level demanded.

So, let's return to the question about the nature of perfection activity indices when learning and turning from one strategy into another.

Let's consider a very simple example. We teach a subject to identify (make a diagnosis) 16 different states of some object. An object is reflected by means of 30 informational elements and each one can have two different states : parameter displaying normally and patologically.

At the first stage of learning the subject evaluates the object's state looking over all 30 elements. The informational contents of signal, i.e. of concrete perceiving state of parameter complex, $I_a=30\log_2 2=30$ bits.

At the second stage of learning the subject groups 30 bits into five functional groups and each one possesses four different states. An informational contents of an analogous signal, concrete state of information display system (DIS), $I_b=5\log_2 4=10$ bits.

Finally, at the third stage the subject learns to identify the state of all the combinations from 30 informational elements as a whole, and for all that the general number of the states of this combination, simultaneously distinguished, is 16. Then the informational contents of signal $I_c=\log_2 16 = 4$ bits.

When comparing with the wholist estimation of each from 16 states, the abundance of informational contents of the signal at the first stage $I_a/I_c=7{,}5$, and at the second stage $I_b/I_c=2{,}5$.

The strategy of the identification of the object's states, from the informational complex state, used by the subject at the third stage of learning is more specialized - it's more adapted for the identification $Q_c=16$ concrete states by means of their differentiation.

The strategy that is used at the second stage is more universal than the strategy existing at the third stage : it helps to differ $Q_b=2^{10}$ states of object. But the second strategy is more specialized than the first : it helps to differ 2^{20} times less states than by means of the first strategy $Q_a=2^{30}$. One can see from it that for relatively minimal decrease of the situations identification time we pay a significant limitation of freedom degrees number : $t_a/t_b=1{,}33$; $Q_a/Q_b=2^{20}$; $t_a/t_c=2{,}35$; $Q_a/Q_c=2^{26}$.

And now the implication follows that each following strategy as a rule, improves the decreasing of fulfilling tasks complexity at the expense of its more specialization as compared with the preceding strategy.

If we return to the data about the solving problems in mnemoscheme, then graphically this phenomenon can be expressed by the fact that the width of the B strategy curve (an interval of values of n or other psychological factors of problem solving complexity (K_{JR}) in which the B strategy helps to manage with the problem) at the Fig. 11 is more narrow than the width of the A strategy. Accordingly, the interval of indices n, in which the strategy C can be realized, is less than the intervals for the strategies A and B.

Stated otherwise, although the strategy A has an optimum which is worse than those of the B and C strategies ($t^a_{min} > t^b_{min} > t^c_{min}$), and simultaneously the B strategy can, even with difficulty, be realized when n moves aside significantly from n^A_{opt} - an optimal index of this factor (under the circumstances of $t > t^a_{min.}$). In experiments on the solving operative problems in mnemoscheme, is found that this interval for the A strategy is Δn^A and for the C' strategy is $\Delta n^C << \Delta n^A$. The curves on the left side we name characterization curves of strategies.

On the right side of Fig. 11 is shown the constructing of the learning curve, i.e. of the change of the meantime of the solving problems t in mnemoscheme in the current of the time of learning T.

We transfer into the right side of Fig. 11 the indices of t_{an} (the time of solving at the beginning of learning), $t_{A\ min}$, t_{AB}, t_{BC}, $t_{C\ min}$.

The number of the informational elements in the given case is looked through as a generalized psychological factor of the complexity [21, 31]. n is a real index of the psychological factor of complexity, n_{opt} is its theoretically optimal index for the concrete strategy (i=A, B, C); n^{AB} and n^{BC} are the indices of the psychological factor of complexity which are common for both strategies. Pointers in the left side of Fig. 11 show the direction of the $t(K_{jp})$ change during the learning. Turning from one strategy into another in the given case occurs when $n^{AB} < n^A_{opt}$ and $n^{BC} < n^B_{opt}$, i.e. when the real index of the psychological factor of complexity decreases as compared with its optimal index for given strategy. Consequently, $t_{AB} > t_{A\ min}$ and $t_{BC} > t_{B\ min}$. This fact does explain the emergency of humps when turning from the A strategy into the B strategy and from the B strategy into the C strategy, with the maximum at the moments T_{AB} and T_{BC}.

If we would take as the criterion of complexity rather another index than the time, for instance, the effectivity or the speed of the problem solving which is adverse to the solution time, then the graphics $Q(K_{jR})$ and $Q(T)$

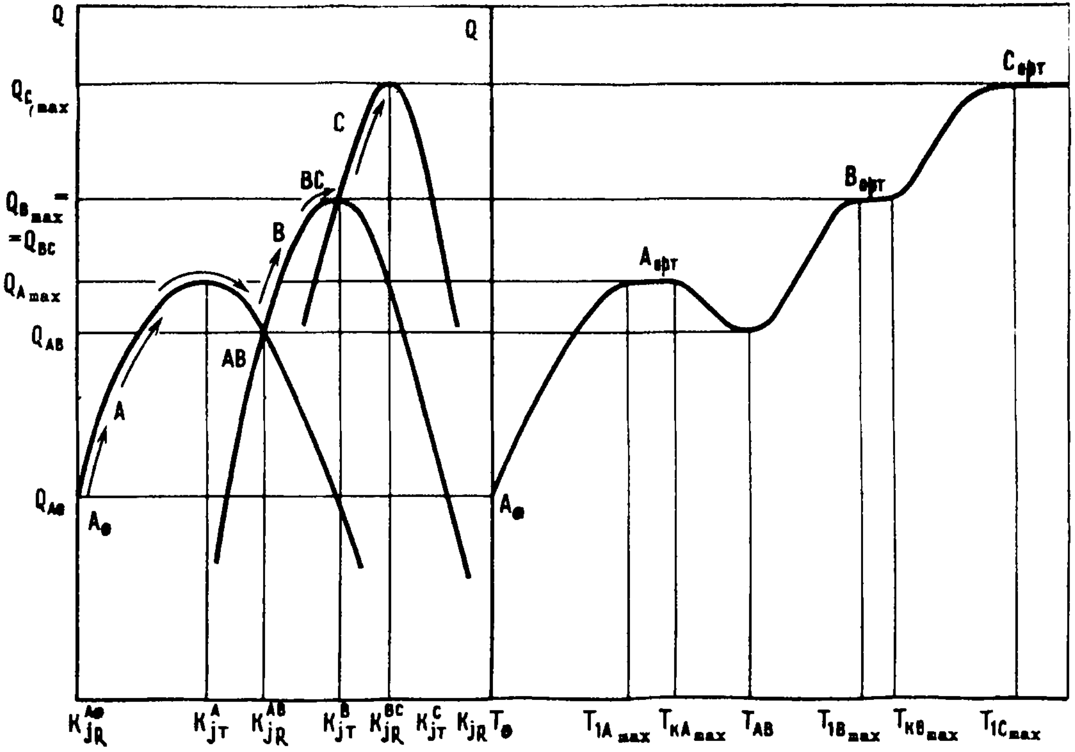

Q
Q
$Q_{C,max}$
C
BC
$Q_{B_{max}} = Q_{BC}$
B
$Q_{A_{max}}$
Q_{AB}
AB
A
Q_{A0}
A_0
A_0
A_{opt}
B_{opt}
C_{opt}
K_{jR}^{A0}
K_{jT}^{A}
K_{jR}^{AB}
K_{jT}^{B}
K_{jR}^{BC}
K_{jT}^{C}
K_{jR}
T_0
$T_{1A_{max}}$
$T_{KA_{max}}$
T_{AB}
$T_{1B_{max}}$
$T_{KB_{max}}$
$T_{1C_{max}}$

Figure 12

Constructing of learning curve (right) with characterization curves of strategies A, B, C for criterion of effectiveness Q.

would look as shown in Fig. 12, where the case when K_{jR} increases in the process of strategy evolution and transformation during the learning, is presented.

The different relative positions of characterization curves of strategies give us different transformation phases on the learning curve. Optimums of the A and B strategies (see Fig. 11, 12 - left) are far each from another and their common point AB is very low. That is because we find downhill the phase of A to B transformation at the learning curve (see Fig. 11, 12 - right) ; B and C strategies are closer and transformation phase has only one plateau. Later we shall show particular cases where transformations may be almost exponential monotonous.

It's well known that the solving of the complex problems, particularly, connected with the operative control of objects, is characterized by the presence of a wide range of variants. There can be, theoretically, an infinite number of solving strategies.

Fig. 13 shows a certain set of really possible strategies (this set is practically limited by the technical devices, abilities of different people to relate one to another, to translate the experience, to standardize it and so on).

If we draw our attention to the decreasing of the learning then in spite of turning the strategy A into B-C-D-E-F, we would realize faster turning of the A strategy directly into D-F, although, it's connected with great provisional difficulties and with the decreasing of Q effectivity during the periods T^{*}_{AD} and T^{*}_{DF}. When this turning exists, the level of the index of effectivity criterion Q_{Fopt} would achieve T_{Fopt} before the sequential turning from A into B-...-F.

The analysis shows that the learning process, which is reflected by the curve Q(T) may proceed monotonously, without "humps", if the curve of the following strategy $q_i(K_{jp})$ crosses the curve of the preceding strategy $Q_{i-1}(K_{jp})$ near by its optimum. The gradual "peaceful growing" of the new strategy in the old one is proceeding.

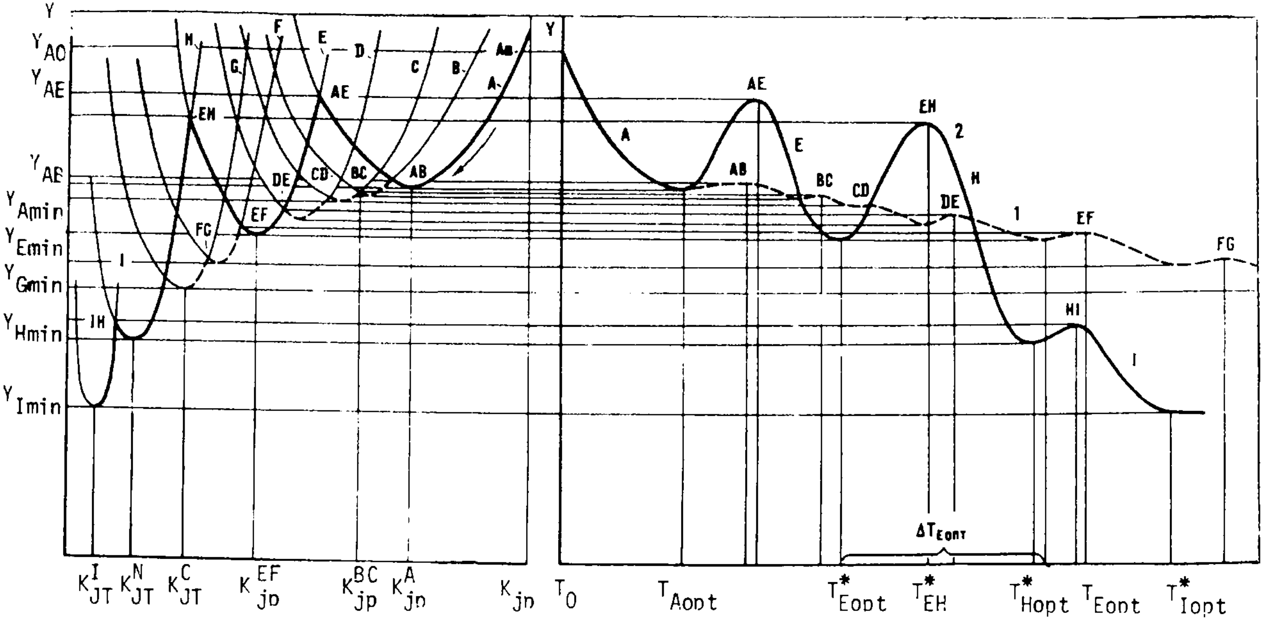


Figure 13

Constructing of different learning processes with different transformation of strategies.

The educational experiments aimed at the experimental verification of the hypothesis about the rise of "humps" during the turning from one strategy into another and about the dependence of the "humps" height upon the distance between the strategies, namely, between the optimal qualities of the leading factors of complexity were carried out and the different types of problems and the technical learning devices were used in it.

Jointly with Y.M. Magazanik and M.V. Rybalchenko the comparison of the learning processes in operators of power plants using usual mnemoschemes (1), mnemoschemes with the elucidation of the actual contours connected with the problems (2) and the command and informational mnemoschemes (3), which reflect visually the algorithm of the problem solving were also carried out [31] The experiments have shown that the regulation of the visual material at the stage of mnemoscheme design brings the number of the strategies successively acquired by the students down, but when displaying the algorithms, only the evolution of the unique strategy C_3 strictly determined by the structure of the informational system occurs, and that's why the curve 3 in Fig. 14 has a monotonous asymptotic character. Curve 2 reflects the sequential acquiring of two strategies B_2 and C_2, and curve 1, of all the main strategies A_1, B_1 and C_1.

A_1 includes the reviewing of all the informational field and the distinguishing of the part of information, which related to the concrete problem B_1 permits to begin the processing of the relevant information at once, because the contours actual for the problem are elucidated in the mnemoscheme. Strategy C_3 is directly connected with the forming of the solving.

The analysis of non-monotonous curves of learning shows that absolute indices of the complexity criterion (for instance, the solving time) is not enough to evaluate the level of man's learning to solve problems. Let's propose that the studentoperator when working with the usual mnemoscheme (Fig. 14, curve 1) has shown in the experiment t=132 sec. Let's also draw in Fig. 14 a horizontal line, which corresponds to this index and then we conclude that such an index of solving time may be achieved under the circumstances of five qualitatively different ways of solving. The point a corresponds to the A strategy plateau where the subject safely possesses this strategy. The experiments have shown that if the student is at the

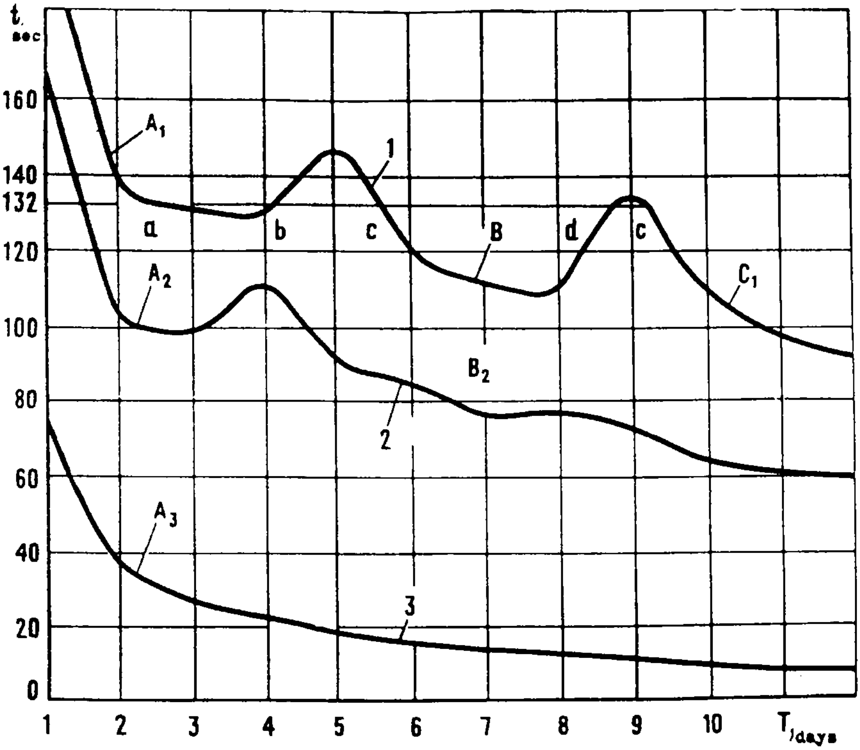


Figure 14

Experimental learning curves.

optimal level in one of the strategies (in the given case of the A strategy), then the structure and indices of his solving operative problems, belonging to all types researched (thinking, perceptual and senso-motor), are maximally stable to the influence of stress factors and minimally dependent on the human's psychophysiological state.

The point b in Fig. 14 corresponds to the identical index t=132 sec., but it lies at the beginning of the turning from the A_1 strategy into the B_1 strategy, and this fact is connected with the deviation from the solving structure optimal for the A strategy and with the partial transformation of this structure. The dispersion of the solving indices in this point, as also in the C point, - in the abrupt loss of the B_1 strategy perfection - is more than in the point a.

The indices in the points b and c are more exposed to the influence of stress factors which affect the man when he solves the problem (the conditions of exams were used as such factors in experiments with the students).

The points d and e lie at the ridge of turning from the B_1 strategy into the C_1 strategy. There is a maximal dispersion of indices.

The analysis of the learning curve 1 in Fig. 14 shows that the drawing attention at the absolute index of complexity criterion is not enough to assert about the level of the subject's training; it's a necessary qualitative and quantitative analysis of the strategy with the help of which this index of criterion is achieved.

Under the circumstances of t=132 sec., the learning can be stopped only in one from five points, which corresponds in curve 1 to this index, namely, in the point a, when the student acquires safely the A_1 strategy, although it's minimally perspective from the possible set A_1, B_1, C_1.

To brake the learning in the point c isn't an expedient thing, because then the small interval of learning (nearly two days) is enough for significant decreasing and stabilization of the t index. If the learning in b or still worse in point d brakes, the student appears to be in the very difficult situation : he does safely possess no one strategy.

We should account the fact that $t_a = t_b = t_c = t_d = t_e$ not only during the organization of the learning process, but also during the solving of the problem of the transfer of the laboratory results into the practice of the real working activity organization.

The method of planning experiments based on the models of R. Bush-F. Mosteller, is described some pages later. Its significant element was the comparison between the solving indices of identical problems in experienced operators (in the control experiment with one of the types of tools or of the job conditions) and in subjects who give the indices for comparative evaluation of different types. Returning to Fig. 14 and proposing that the experienced operators show that during the problem solving $t_{op} = 132$ sec. and it coincided with the indices of subjects $t_{sub} = 132$ sec., then from this coincidence $t_{op} = t_{sub}$ there is not to conclude the correspondence between the solving identical problems by the subjects and experienced operators, because the strategies of solving may be significantly different.

Thus, also, during the organization of a learning process, choosing of its longitude, finishing level of indices is needed, and during the solving question about the practical usefulness of the psychological experiments together with the absolute indices must be compared the strategies of solving with the help of which the indices given are achieved.

The main attention must be drawn at the analysis of the strategy transformation as the most complex and unsteady zones during the learning.

It's necessary to say that the concepts of stability and unstability are certainly relative; they must be added by the quantitative values. In the experiments described, it is found that at the analogous phases of acquiring different strategies, the dispersion is less when using a more perfected (precisely, more specialized) strategy.

Moreover, the significant unevenness of solving stability indices when acquiring one strategy is found. For instance, when learning to solve problems in the command and informational mnemoscheme, and this learning is characterized by the strategy stability, there are some zones with the decreasing dispersion of indices.

It gives us a possibility to make a proposal about few levels of detality of learning processes research. When we study the process of the perfection the ways of achieving the certain class of aims and of the solving the corresponding problems at all, the possible strategies may be united by the more general concept of solving policy. At this level of analysis of the policy, the transformations between the strategies may be reviewed as monotonous. Accordingly, namely, this is proposed by the authors of "Introduction to ergonomics" who work out a "macro-approach" to the learning processes analysis.

From the other side, not only the research of the solving policies, but also of strategies must become detalized when researching individually adapted methods of learning, controlling the experiments and so on. In this case, the strategy must be segregated into tacts - more narrow classes of the solving ways.

In this case, it's possible sometimes to find the "humps" even in such smooth, monotonous curves as C_3 in Fig. 14. In their turn, tactics are divided into concrete ways and acts of solving and also may be presented as the curves with humps. It seems to be explained by the general regularity of the oscillations of the solving processes as such.
U - oid curves analogous to the curves in the left part of Fig. 13, qualitatively reflect the dynamics of policies in a whole, strategies, tactics, concrete solving and even the single step in the solving. If Q is interpreted as the success probability in each step of learning, then the curve at the right side would show the track of oscillations - search steps around the maximal success probability. Such a track would show the regular (but not trial and error) search of solving based on the using of policies, strategies, tactics, ways of solving worked out by the given individual and his predecessors and analogous to met in the given case. Each step of solving may be connected with the optimal index of a particular psychological factor of complexity. It's necessary to emphasize the conventionality of the presentations of dependences $t(K_{jp})$, $Q(K_{jp})$ at the

plane; everywhere we must look at the few graphics in connection with the number of psychological factors of complexity simultaneously. In [22,31] we reviewed 15 different K_{jn} in the case of the analysis of operative problems solving. It is necessary to create special methods of the graphical displaying of such multi-factor processes of the optimization of the informational base of solving, in which the solving becomes accessible to the concrete individual in the concrete experience (when we realize the individual level of analysis), to the contingent of specialists or to the "mean" man (when we realize the total level of analysis and experimental data generalization). As we see in Fig. 8, the research of the influence of the technical devices types to the processes of evolution and transformation of the strategies takes an important part for the learning processes optimization.

In the experiments held by our laboratory the learning of three groups of operators was studied (see Fig. 15). The first group (curve 1) solved the problems in the text instructions, the second group in the command and informational mnemoschemes (curve 2). The third group after few lessons, connected with the text instructions, turned into the solving with the help of command and informational mnemoschemes. Although the solving of operative problems with the help of such mnemoschemes is significantly easier than in the text instructions, the turning from instructions into mnemoschemes is connected with significant provisional increasing of the solving complexity, and then this complexity rapidly decreases (see Fig. 15, curve 3).

Y.A. Gvozdev's experiment connected with teaching the group of students to read the operative information quickly using the metronome which set the rhythm of reading and pacer, which helped to regulate the speed of the transference of text projected to the screen, confirmed the regularities of evolution and transformation found (see Fig. 16).

It is necessary to point out, that the phenomena analogous to such which are observed when transforming strategies during the learning of the man to the operative perception and processing of the information, are characteristical for a great number of other processes of human's adaptation to the external conditions.

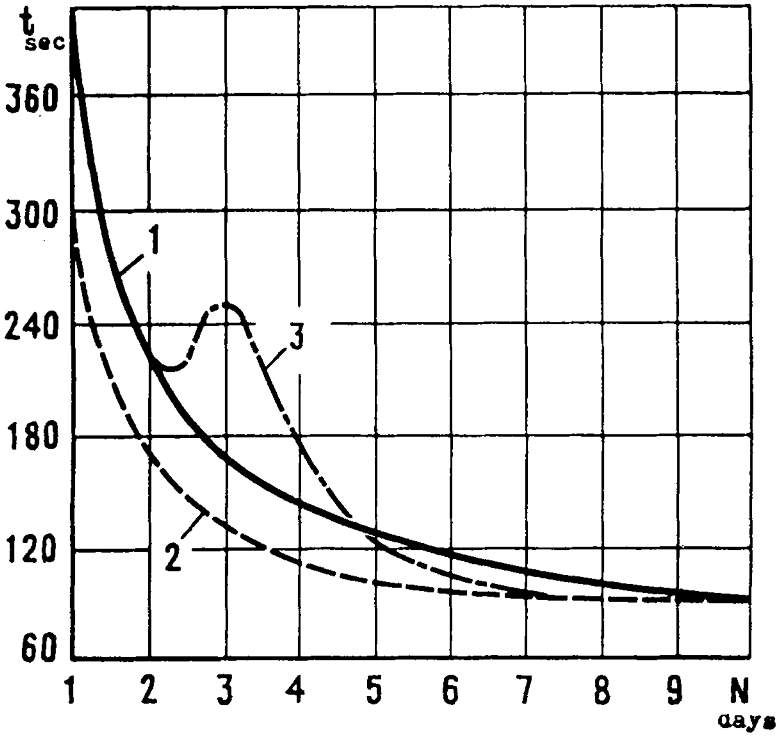


Figure 15

Experimental learning curves on the operative problem solving with mnemonic schemes as the displays.

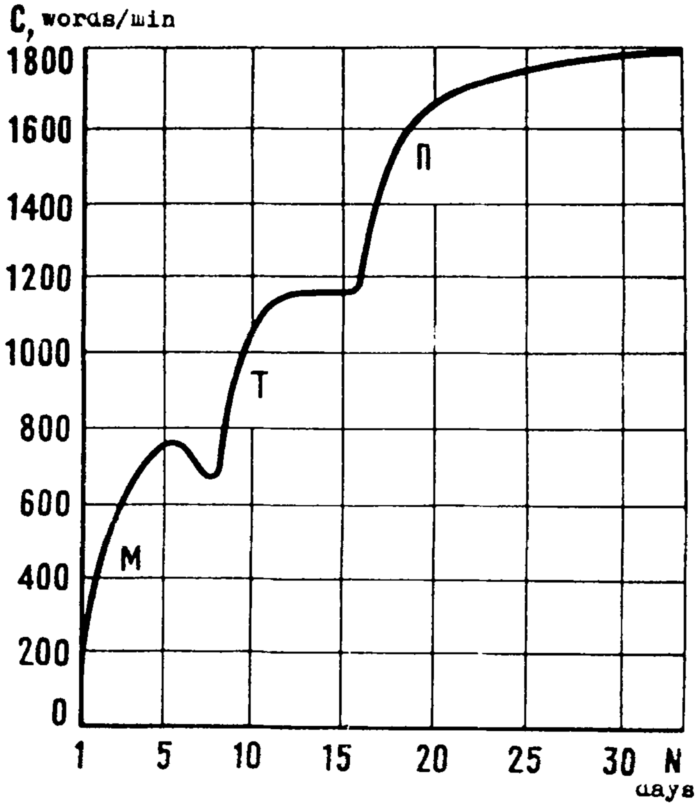


Figure 16

Experimental learning curve on the speedreading.

C : read words per one minute

N : days of learning

Thus, for instance, B.F. Lomov showed the process of dark adaptation of the human's visual analizer when the colour characteristics of light are different. F. Sneider and N. Pronco taught men to group game cards accordingly with the different signs and put on subjects glasses reversing the image during the period from the 17th to 45th experiment. After the glasses were taken off, the temporal decreasing of indices was observed, identically to the fact of significant decreasing of operator's work when he turns from the complex text instruction into more simple command and informational mnemoschemes.

3. CONCLUSIONS

Thus, the systematic studies of characterization equations, curves of strategies and of evolutions and transformations of strategies during the learning were the fundamentals of a new transformation theory of learning processes. The theory has a theoretical and practical significance, particularly, for the solving of the problem of the optimal control of the learning process, and for more perfect experimental and analytical models of this process building.

The process of learning may be described by the traditional monotonous models only in particular cases, at the stage of the concrete strategies evolution or when the subject consequently acquires nearly lying strategies and also when it's averaging the great arrays of experimental data.

The individualization of the methods and means of the operator's learning which becomes wide-spread nowadays makes quite actual the development of a "micro-approach" to the study of learning processes together with the traditional "macro-approach".

Monotonous and non-monotonous (with intermediate plateaus and "humps") forms of curves are significantly dependent on the level of psychological analysis of the learning processes. Practically, it is always possible to find a quite high level of generalization, in which the learning process would be approximated by the smooth monotonous curve. From the other side, it's possible to achieve such a level of analysis and such detalization degree as to the learning processes dynamics, in which the transitional processes which present a special interest, would be found.

For the analysis of the stages of strategies transformation, the new types of models, particularly, grapho-analytical models must be worked out. The main nowadays is the accumulation of the experimental material without the ancient traditions of "smoothing" which mask the most interesting psychological pecularities of the learning processes. The using of the system approach proposes to study it in the different levels: from the total up to individual and operative, from the policy up to single step within the concrete act of solving, and also from the viewpoints of complexity, stability and effectivity of the solutions.

The development of the "micro-approach" to the analysis of the learning processes and transformation theory of learning give us a possibility to use them in a broader class of cognitive processes as a method of analysis of cognitive dynamics in the scientific and technical creativity, technological processes and so on.

REFERENCES

[1] Atkinson, R.C., Crothers, E.J., A comparison of paired-associate learning models having different acquisition and retention axioms, J. of Mathematical Psychology, 1 (1964), 285-315.

[2] Bower, G.H., Application of a model to paired-associate learning, Psychometrika, 26 (1961), 255-280.

[3] Bruner, J.S., Goodnow, J.J., Austin, G.A., A study of thinking, New York, Wiley (1956).

[4] Bush, R.R., Mosteller, F., Stochastic models for learning, New York, Wiley (1955).

[5] Estes, W.K., Toward a statistical theory of learning, Psychological Review, 57 (1950), 97-107.

[6] Estes, W.K., The statistical approach to learning theory, in S. Koch (Ed.), Psychology: a study of a science (V.2), New York, McGraw-Hill, (1959).

[7] Greeno, J.G., Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters, J. of Mathematical Psychology, 4 (1967), 430-472.

[8] Hull, C.L., Principles of behavior: an introduction to behavior theory, New York, Appleton-Century-Crofts (1943).

[9] Luce, R.D., Individual choice behavior: a theoretical analysis, New York, Wiley (1959).

[10] Postman, L., Underwood, B.J., Critical issues in interference theory, Memory and cognition, 1 (1973), 19-40.

[11] Restle, F., The selection of strategies in cue learning, Psychological Review, 69 (1962), 329-343.

[12] Skinner, B.F., The behavior of organisms: an experimental analysis, New York, Appleton-Century-Crofts (1938).

[13] Thorndike, E.L., The psychology of arithmetic, New York, MacMillan, (1922).

[14] Tolman, E.C., Purposive behavior in animals and men, New York, Century (1922).

[15] Woodworth, R.S., Experimental psychology, New York, Holt (1938).

[16] Chapanis, A., Research techniques in human engineering, Baltimore (1959).

[17] Ebbinghaus, H., Ueber das Gedachtnis, Leip-Duncker (1885).

[18] Thorndike, E.L., Animal intelligence: an experimental study of the associative processes in animals, Psychol. Monogr., V.2, N°8 (1898).

[19] Vincent, S.B., The function of the vibrissal in the behavior of the white rats, Behavior. Monogr. V.5 (1912).

[20] Woodworth and Schosberg's Experimental Psychology, ed. Holt, Rinehart and Winston, New York.

[21] Lomov, B.F., Man and technique, Moscow (1966).

[22] Lomov, B.F., Venda, V.F., Human factors: problems of adapting systems for the interaction of information to the individual: the theory of hybrid intelligence, in Proc. of the Human Factors Society, 21st Annual Meeting, San Francisco (1977).

[23] Lomov, B.F., Venda, V.F., Methodological principles of synthesis of hybrid intelligence systems: Proc. of the Intern. Conf. on Cybernetics and Society, Tokyo (1978).

[24] Sheridan, T.B., Ferrell, W.R., Man-Machine systems: information, control and decision models of human performance, The MIT Press (1974).

[25] Venda, V.F., Information display systems: Ergonomics and design aspects, Moscow (1969). (Russian)

[26] Venda, V.F., The perspectives of development of the psychological theory of learning, Psychological Magazine, 4 (1980). (Russian)

[27] Venda, V.F., Psychological analysis of learning processes dynamics, Paper on USSR/USA and USSR/Finland symposiums on psychology of activity, Moscow-Helsinki (1980).

[28] Venda, V.F., Voies nouvelles pour une théorie de l'apprentissage. Présent et futur de la psychologie du travail, Paris (1980), 507-517.

[29] Venda, V.F., Lomov, B.F., Human factors leading to engineering safety systems. Hazard Prevention, V.15, [illegible], Los Angeles (1980).

[30] Oshanin, D.A., Venda, V.F., Wege zur Erholung des Arbeitseffekts des Operateurs in Systemen "Mensch-Automat", Probleme und Ergebnisse der Psychologie, Berlin (1962).

[31] Venda, V.F., Engineering psychology and synthesis of information display systems, Moscow (2nd ed. (1981)) (1975). (Russian)

[32] Venda, V.F., Ergonomic problems of individual adaptation of operator's work means, Ergonomics, 6 (1976), (Paper presented in the VI Congress of the IEA).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

A PROBABILISTIC CHOICE MODEL ADAPTED FOR THE ANGLE PERCEPTION EXPERIMENT

N.D. Verhelst
State University of Utrecht
The Netherlands

A probabilistic model is developed that is adapted to the data collection method known as "pick any out of n", under the condition that statistical replications of the choices are possible. The psychophysical problem of systematic errors in angle reproduction is chosen as a typical example. Further, it is shown under which conditions the model is not identifiable and it is pointed out that the model in its most general form, although identifiable, may be ill conditioned because the parameter estimation is akin to the problem of estimating sums of exponential functions.

1. INTRODUCTION

In the problem area of scale construction, an important distinction can be made between scales constructed from data which reflect a dominance relation between individual and stimulus on the one hand, and scales constructed from data reflecting a proximity relation between individual and stimulus. This distinction corresponds to Coombs' distinction between QIIa and QIIb data respectively (Coombs (1964)). In the framework of latent trait theory, it can be argued that these two kinds of data - as far as binary data are considered - are reflected in models with monotone and single-peaked item characteristic functions, a distinction labeled by Torgerson (1958) as models with monotone vs. nonmonotone items. In this paper we will be concerned with the construction of a probabilistic model for scaling binary proximity data.

A typical application is the scaling of objects from free binary choice data, i.e., the subject is free to choose any out of a number of objects (the so-called pick any/n data gathering strategy).

In many scaling applications however, especially where antitudes, attitudes or preferences are studied, a double problem usually must be solved: the stimuli as well as the individuals must be scaled. In the context of preference research, the scale value attached to the individual is called his ideal point, a term which we will use generically in a broader context.

If there is only one response per stimulus-person combination available, the aggregation of data (over persons for example) is made difficult, because each individual will enter the model with a new parameter - his ideal point - and these parameters will act as nuisance parameters in th estimation of the scale values of the stimuli. This problem is well known in latent trait theory (Fischer (1974)). Therefore it is interesting to consider situations where the researcher has the opportunity to accumulate information without the introduction of extra nuisance parameters, e.g., by having pure replications of the data. Therefore, we will assume that the basic data we have at our disposal consist of an individual x stimulus x replication array of binary data, and that in the model to be developed, parameters are associated with stimuli and individuals, but not with the replications.

In order to make the choice of the model plausible, we will develop it in the light of a psychophysical problem to be introduced in the next section.

2. A PSYCHOPHYSICAL PROBLEM

An old, but up to now unsolved problem is given by the observation that, when subjects are asked to reproduce an acute angle - either by memory or by simultaneous vision of the standard and his own reproduction - by adding a line to a given horizontal line, the reproduced angle tends to be wider than the standard, whereas the reverse effect was obtained with obtuse standards (Jastrow (1892)). The suggested explanation of a perceptual over- or underestimate of the standard rests on a logical flaw in the argument, because, as was pointed out by Fisher (1969), the same perceptual distortion should apply to the reproduced angle, so that there is no compelling reason why they should differ. MacRae and Loh (1981) tried to evaluate the possible effects of the horizontal-vertical illusion (vertical distances appear bigger than horizontal ones), the direction of the initial adjustments (they used an adjustable angle with fixed bisector; its starting position was either much smaller or much bigger than the width of the standard), and possibly a statistical artifact (the distribution of reproduced angles being skew in opposite directions for acute and obtuse angles; the use of the arithmetic mean as sample statistic might explain the problem). Their findings were generally negative: control over the three likely biasses could not take away the systematic effect in the reproduced angles. Thus, it appears that the problem is rather tough and complicated and that the effect might be caused by multiple sources, e.g. perceptual, motor and decisional processes. In order to disentangle the problem it seems wise to do experiments where the motor component is excluded, and to construct a formal model such that, in combination with good experimentation, the perceptual and decisional component can be separated, much like in signal detection theory (Green and Swets (1966)).

The typical experiment we have in mind is a simple binary classification task, where the subject is asked to judge if an angle (called 'variable' angle) is equal or not to a given angle (called the 'standard'). In one experimental session - or part of it - the standard is fixed and visible throughout, and a series of variable angles is presented to the subject. The whole experiment consist of a number of these subexperiments, each with a different standard. It is hypothesized that the perceptual processes are controlled by the physical characteristics of the stimuli and the general visual context of the experiment, whereas the decision processes involved are under the control of instructions, reward, etc.

It might be suggested that one could equally well ask the subject to judge the direction of the unequality, so that with the same experimental effort, much more information can be obtained. Undoubtedly, this is true, but our main aim is to construct a formal model for cases where such directional information is unavailable. In the case of the angle problem at hand, one could use this richer information to test the more general model. Within the framework of this paper however, we will not consider this possibility in detail.

To keep the use of concepts clear, the correspondence between the generic concepts used in the introduction and the terminology used in the description of the model is listed in table 1.

generic	experimental
stimuli	variable angles
individuals	standard angles
replications	one or several persons at one or more occasion(s)

Table 1

Correspondence between generic concepts and terminology used in the experiment.

3. THE MODEL

We will use the following conventions for the notation.

s : number of standard angles used in the experiment.

j : general index to indicate standard angles.

k : number of variable angles in case this number is constant for all standards.

k_j: number of variable angles in the subexperiment with the jth standard.

i : general index to indicate variable angles.

n : number of replications per standard-variable combination. It is assumed that this number is constant throughout.

v : general index of a replication.

x : width of an angle, expressed in radians.

The model we will discuss is intended to be valid for all angles in a certain open interval (a,b), e.g. $(0, \pi/2)$. The range of the validity is an empirical problem.

The response of the subject in the experiment will be considered as a random variable, defined by

$$A_{vij} = \begin{cases} 1 & \text{if at replication } v \text{ variable angle } i \text{ is judged to be} \\ & \text{equal to standard angle } j. \\ 0 & \text{otherwise.} \end{cases}$$

And we define

$$f_{vij} = \text{Prob}(A_{vij} = 1).$$

The axioms of the model are:

AXIOM 1 : Independence

For all v, i and j the random variables A_{vij} are mutually independent. This axiom amounts to the assumption of experimental independence between subjects and the so-called local stochastic independence within subjects.

For the next axioms, we assume the existence of an unbounded monotone increasing function σ from (a,b) onto the reals:

$$\sigma : (a,b) \to \mathbb{R}$$

and such that

$$\lim_{x \to a} \sigma(x) = -\infty \text{ and } \lim_{x \to b} \sigma(x) = \infty .$$

For conciseness in the notation we will write for all angles x_h, σ_h instead of $\sigma(x_h)$. The range of the function σ will be called the (unidimensional) latent continuum. All the following axioms will be expressed in terms of this latent continuum. The function σ will be called the psychophysical function.

AXIOM 2 : Acceptance region

As a consequence of the experimental procedure and the instructions, an interval (ζ_j, ξ_j) is created on the latent continuum.

This interval will serve as the criterion used by the subject in making his judgments of equality or unequality. It is hypothesized that the width of this interval can be manipulated by instructions and rewards. Notice however, that the validity of this psychological hypothesis is not critical for the validity of the model as a formal description of observations, giving room to test hypotheses about psychological processes.

AXIOM 3 : Acceptance processes

Upon the presentation of a variable stimulus i, two independent (latent) processes take place, which we will denote as the lefthand process and the righthand process. Roughly spoken, the lefthand process checks whether the variable angle is not too small to be acceptable, and the righthand process checks whether it is not too big. We therefore define two latent random variables $A^{(\ell)}_{vij}$ and $A^{(r)}_{vij}$:

$$A^{(\ell)}_{vij} = \begin{cases} 1 & \text{if variable angle i is accepted by subject v in the lefthand latent process.} \\ 0 & \text{otherwise.} \end{cases}$$

$A_{vij}^{(r)}$ is defined analogously for the righthand latent process.

We define further:

$f_{vij}^{(\ell)} = \text{Prob}(A_{vij}^{(\ell)} = 1)$ and $f_{vij}^{(r)} = \text{Prob}(A_{vij}^{(r)} = 1)$.

Now, the axiom on the acceptance processes can be stated as follows:

3.a $A_{vij}^{(\ell)}$ and $A_{vij}^{(r)}$ are stochastically independent

3.b $f_{vij}^{(r)}$ is a monotone increasing function of $\xi_j - \sigma_i$

3.c $f_{vij}^{(\ell)}$ is a monotone decreasing function of $\xi_j - \sigma_i$.

It should be stressed that Axiom 3 is concerned with latent processes, which are unobservable, and that it does not follow from the axiom that always two independent psychological processes are assumed. This will be clear from the example following axiom 4.

AXIOM 4 : Response axiom

The observable response variable A_{vij} is defined to be

$$A_{vij} = A_{vij}^{(\ell)} A_{vij}^{(r)} .$$

Because of axiom 3a it follows immediately that

$$f_{vij} = f_{vij}^{(\ell)} f_{vij}^{(r)} .$$

Now, as an example consider the psychological theory which says that the two latent processes are not independent, but that the subject initiates one latent process, e.g. the lefthand one, and that he only initiates the righthand one if necessary, i.e. if the variable angle is not judged too small. Furthermore, there is some probability g that the process he begins with is the lefthand one. If we denote the probability that the angle is not judged too small by the lefthand process by p, and the probability that the angle is not judged too big by the righthand process by q, and assuming that an angle that is neither judged too small nor too big, is judged to be equal, then

$$\begin{aligned} \text{prob(equal judgment)} &= g\,p\,q + (1-g)p\,q \\ &= p\,q \end{aligned}$$

which is formally the same as stated immediately following axiom 4. That means that, as far as the observable binary responses are concerned, the psychological theory sketched is formally equivalent with axioms 3 and 4.

The model is not completely specified by the foregoing axioms; especially the class of functions $f_{vij}^{(\ell)}$ and $f_{vij}^{(r)}$ has infinitely many members. We choose the logistic function because it has proven to be very useful in other contexts. Thus we define the model in its most general form as

$$f_{vij}^{(r)} = \frac{\exp\{\alpha_{ij}(\xi_j-\sigma_i)\}}{1 + \exp\{\alpha_{ij}(\xi_j-\sigma_i)\}} \qquad (\alpha_{ij}>0) \tag{1}$$

$$f_{vij}^{(\ell)} = \frac{1}{1 + \exp\{\beta_{ij}(\zeta_j-\sigma_i)\}} \qquad (\beta_{ij}>0). \tag{2}$$

However, the model as specified in (1) and (2) is not identifiable because with s standards and k variable angles, and assuming that all combinations are tested, the model has $sk + k + 2s$ parameters while there are only sk cells with information about the parameters since the replications are statistically equivalent. Therefore we need further structural constraints on the parameters. To simplify the exposition, we will assume that

for all i, $(i=1,\ldots,k)$, $\alpha_{ij} = \alpha_j$ and $\beta_{ij} = \beta_j$ $(j=1,\ldots,s)$.

It is clear that other constraints could have been chosen and this problem will be discussed further in the section about identifiability. With this simplification the model can be stated as:

$$f_{vij}^{(r)} = \frac{\exp[\alpha_j(\xi_j-\sigma_i)]}{1 + \exp[\alpha_j(\xi_j-\sigma_i)]} \qquad (\alpha_j>0) \tag{3}$$

$$f_{vij}^{(\ell)} = \frac{1}{1 + \exp[\beta_j(\zeta_j-\sigma_i)]} \qquad (\beta_j>0). \tag{4}$$

It is clear that in (3) and (4) a multiplication of α by a positive constant can be neutralised by the multiplication of β and the division of ξ, ζ and σ by this same constant. Analogously, the addition of an arbitrary constant to σ can be neutralised by the subtraction of the same constant from ζ and ξ. In other words, we are free to choose the unit and origin of the latent continuum.

As to the interpretation of the modelparameters, it may be observed that when α and β both tend to infinity, the product of (3) and (4) has a step function as it limit, whose value is 1 in the open interval (ζ,ξ) and zero outside it. That means that once the boundaries of the acceptance region are set, all uncertainty about the responses has disappeared. If however, α and β become small, f_{vij} is a smooth single peaked function, whose maximum value is less than unity. That means that the decision criteria become less sharp, and it is hypothesized that this is a consequence of noise in the perceptual processes. However, it should be stressed again, that the correspondence between perceptual processes and the value of α and β on the one hand, and between decisional processes and the width of the acceptance region on the other hand constitutes an empirical problem. The model defined by (3) and (4) offers a formal means to distinguish between them. It may be useful to observe that the model does not parametrize the ideal point, i.e., there is no scale value associated directly to the standard angle x_j. The scale value of x_j may be estimated indirectly by using a variable stimulus x_i such that $x_i = x_j$.

Suppose we observe the same phenomenon in the classification experiment as did Jastrow in the reproduction experiment, i.e., with an acute standard angle x, the variable angle $x + x'$ ($x' > 0$) is judged more often equal to the standard than the variable angle $x - x'$, then there are at least three different reasons possible to explain this phenomenon. We will consider them in turn.

Case 1. $\alpha_j = \beta_j$ and $(\xi_j - \sigma_j) = (\sigma_j - \zeta_j)$.

In this case the function f_{vij}, considered as a function of $\sigma(x_i)$ is symmetric with regard to the midpoint $(\zeta_j + \xi_j)/2$. But the psychophysical function σ may be nonlinear in such a manner that

$$\sigma^{-1}(\xi_j) - \sigma^{-1}(\sigma_j) > \sigma^{-1}(\sigma_j) - \sigma^{-1}(\zeta_j).$$

That means that the function $f_{vij}\sigma(x_i)$ considered as a function of x_i has become asymmetrical. So the reason for the asymmetry in the data is explained by the non-linearity of the psychophysical function σ.

Case 2. $\alpha_j = \beta_j$ and the psychophysical function is linear.

The observed asymmetry in the data can be explained by the inequality

$$(\xi_j - \sigma_j) > (\sigma_j - \zeta_j).$$

If this is the case, this asymmetric acceptance region may be made symmetric by manipulation of the instruction and/or rewards.

Case 3. $\xi_j - \sigma_j = \sigma_j - \zeta_j$ and the psychophysical function is linear.
Here the observed asymmetry is due to the difference between α_j and β_j and may reflect genuine perceptual differences.
It can easily be shown that the predicted asymmetry occurs when $\alpha_j < \beta_j$.

4. PARAMETER ESTIMATION AND IDENTIFIABILITY

We will first consider an experiment with one standard, henceforth referred to as the standard experiment. To have a convenient form for the estimates, we define (dropping the now superfluous indices v and j)

$$\beta = 1$$
$$\xi = 0$$
$$e^{-\sigma_i} = \varepsilon_i \quad (i=1,\dots,k)$$
$$e^{\zeta} = \theta.$$

The item-characteristic function, i.e. the product of (3) and (4) can now be written as

$$f_i = \frac{\varepsilon_i^{\alpha}}{(1 + \varepsilon_i^{\alpha})(1 + \theta\varepsilon_i)} \qquad (i=1,\dots,k). \tag{5}$$

Notice that if $\zeta \to -\infty$, the second factor in the denominator of (5) becomes unity, and the resulting model is the simple 2-parameter logistic model (Birnbaum (1968)).

It can readily be seen that the likelihood of the data is given by

$$L = \prod_i^k f_i^{t_i}(1 - f_i)^{n-t_i} \tag{6}$$

where

$$t_i = \sum_v a_{vi}\,. \tag{7}$$

That means that the simple total scores are the minimal sufficient statistics for the whole set of parameters. Now consider a transformation to get the likelihood function (6) in some standard form.

Define

$$z_i = \varepsilon_i^{\alpha} \tag{8}$$

$$y_i = \theta\varepsilon_i\,. \tag{9}$$

The likelihood can now be written as

$$L = \prod_i ((1+z_i)(1+y_i))^{-n} \exp \sum_i (t_i \ln z_i + (n-t_i)\ln(1+y_i+y_iz_i)). \qquad (10)$$

Now define

$$\eta_i = \ln(z_i/(1+y_i+y_iz_i)). \qquad (11)$$

Substituting (11) in (10) gives

$$L = \prod_i (e^{\eta_i} + 1)^{-n} \exp \sum_i t_i\eta_i . \qquad (12)$$

That means (see Andersen (1980)) that the model belongs to the exponential family, with k canonical parameters η_i. It follows from standard theory that the maximum likelihood estimates, if they exist, are unique. Furthermore, the estimates of the canonical parameters are extremely simple: differentiating (12) with respect to η_i and setting zero gives the likelihood equations

$$\frac{\exp(\eta_i)}{1 + \exp(\eta_i)} = t_i/n \qquad (i=1,\ldots,k) \qquad (13)$$

with solution

$$\hat{\eta}_i = \ln\,(t_i/(n-t_i)) \qquad (i=1,\ldots,k) \qquad (14)$$

so that it is readily seen that a solution exists if and only if

$$0 < t_i < n.$$

As to the original parameters, we must say that there are infinitely many solutions, because there are k+2 parameters whereas there are only k sufficient statistics. So, in order the model to be identifiable at least two constraints must be put on the parameter space, for example by fixing two parameters at a given value. Even then, however, the problem is not solved completely, the reason being that the canonical parameters are nonmonotone transformations of the original model parameters. To understand the implication of this, consider the following example: suppose we look for the solution set under the restriction $\alpha = 1$. It follows from (8), (9), (11) and (14) that

$$\frac{\varepsilon_i}{1 + \theta\varepsilon_i + \theta\varepsilon_i^2} - g_i = 0$$

or equivalently,

$$(\theta g_i)\varepsilon_i^2 + (\theta g_i - 1)\varepsilon_i + g_i = 0$$

where

$$g_i = t_i/(n-t_i).$$

By a suitable choice of θ, there are two solutions for each ε. The intuitive reason for this is that (5), as a function of the scale value of the variable angle, is single peaked, so that for each value of ε, say $\hat{\varepsilon}$, there is another value, ε', for which (5) takes the same value. Since the likelihood function is a product of functions like (5), it is clear that the value of L at a local maximum, say $(\hat{\varepsilon}_1,\ldots,\hat{\varepsilon}_i,\ldots,\hat{\varepsilon}_k)$, can be attained at all 2^k points, where any of the $\hat{\varepsilon}$ is replaced by the corresponding ε', and that all those points correspond to local maxima. Since the canonical parameters are unique, we can state that, when α and θ are suitably fixed, there can be up 2^k local maxima.

The way out of this indeterminateness consists of fixing the order of the ε's on the latent continuum: changing an $\hat{\varepsilon}$ into the corresponding ε' necessarily changes the order between two or more parameters, except for the one in the 'middle', i.e. the one with the highest f-value (see (5)). In case of the present experiment where the psychophysical function is monotonic, the order of the ε-parameters constitutes no problem, but in the more general case e.g. the purchasing of goods, finding the order on the latent continuum is one of the major goals of collecting the data.

In summary then, two problems can be discerned in connection with the standard experiment:

(i) the problem of overparametrization, there being more parameters than sufficient statistics, and

(ii) as an immediate consquence of the use of a non-monotonic item-characteristic function, the indeterminateness of any solution unless sufficient ordinal information on the parameters is available.

The way out from these two problems consists of a combination of several standard experiments and capitalizing on the assumption that the psychophysical function is independent of the experimental context. I.e. perform s standard experiments, each with k_j variable stimuli, but chosen in such a way that the number of physically different stimuli, say m, is less than Σk_j.

In such a case the number of minimally sufficient statistics, i.e. the number of canonical parameters is Σk_j, while the number of model parameters can be made smaller.

We can then consider two main extensions of the model, one concerning the discrimination parameters α and β, and one concerning the width of the acceptance interval. For the discrimination parameters consider the following hierarchically ordered submodels.

(1) $\alpha_1 = \ldots = \alpha_j = \ldots = \alpha_s$ and $\beta_1 = \ldots = \beta_j = \ldots = \beta_s = 1$.
That means that one discrimination parameter has to be estimated.

(2) The α's are not constrained to be equal, but the β's are, or the converse. In this case s parameters are to be estimated, the common α or β taken as unity.

(3) No constraints on the discrimination parameters. 2s - 1 parameters must be estimated, one taken as unity.

Crossed with the foregoing cases, we can discern two submodels, concerning the width of the acceptance intervals.

(A) The width of all intervals is constant and unity, leaving s parameters to be estimated, e.g. the lower boundaries ζ_j.

(B) No constraint on the width of the intervals, leaving 2s - 1 parameters to be estimated.

To be identifiable, Σk_j must not be less than the number of free parameters as shown in table 2 for the 6 cases just mentioned.

	1	2	3
A	m+s+1	m+2s	m+3s-1
B	m+2s	m+3s-1	m+4s-2

Table 2

Number of free parameters for several submodels. The variable stimuli always giving m free parameters.

Since the models are hierarchically ordered - all the models being special cases of all other models appearing to the right and/or below their cell in table 2 - they can be tested against each other e.g. by using the likelihood ratio.

The problem of identifiability however, is not necessarily solved because it is not certain that by adding more canonical parameters while keeping the number of modelparameters within the limits given in table 2, the order of the scale values of the variable angles will be unique. To see this, consider the situation where for an experiment with 2 standards and 5 common variable angles, labeled A, B, C, D and E respectively, a local maximum of the likelihood function has been found. Then, for all variable angles, the probability of being judged equal (the product of (3) and (4)) can be calculated, and these function values can be ordered for each of the two standards, thus giving 2 I-scales as defined in unfolding theory. As an example consider the two I-scales to be

I-scale 1 : B C A D E
I-scale 2 : C B D A E.

Then the possible candidates for the order of the stimuli on the latent continuum are a subset of the J-scales compatible with the two I-scales. In this case these J-scales are A B C D E and E A B C D.

The solution of this problem of course is to add more standards in such a way that the I-scales derived determine uniquely the J-scale, and generally this will be accomplished by choosing standards with scale values which are well scattered over the latent continuum, an idea which is very central in unfolding theory (Coombs, chapter 5 (1964)).

It should be clear that the family of models, as summarized in the margins of table 2, does not exhaust the model: as was indicated in section 3, it is possible to conceive of the discrimination parameters α and β, as being a function of the variable angles, or of some specified combination of two parameters, one dependent on the variable angle, the other on the standard.

Although the model as it is defined now, is very flexible and can be adapted and extended in a great number of ways - especially in cases where the control over stimuli, standards as well as variable, causes no problem and where the replications are easily obtained - one should not be too optimistic about the possibilities of testing theories of perception and decision making on one and the same data set, by continuously imposing restriction on the parameter space until the model cracks down. The reason for this is that generally little is known about the power of the tests used. In the case at hand, however, the power of some tests - especially those concerning the equality of the discrimination parameters - will generally be very low as may be seen by the following simplification of the general model for the standard experiment. Some evidence is available in the literature that the psychophysical function for angles is linear (Stevens and Galanter (1957)). So it might seem reasonable to perform the parameter estimation of α, β, ζ and ξ with the scale values as known constants, i.e. $\sigma_i = x_i$.

Defining

$$\theta = e^{\zeta}$$

$$\tau = e^{\xi}$$

$$g_i = t_i(n-t_i) \text{ with } t_i = \sum_v a_{vi}$$

and assuming for the moment that there were 4 variable stimuli in the experiment, then the maximum likelihood estimates are given by the solution of the non-linear system

$$g_i^{-1} = \tau^{\alpha x_i} + \theta^{-\beta x_i} + \tau^{\alpha x_i}\,\theta^{-\beta x_i} \qquad (i=1,2,3,4). \tag{15}$$

But (15) is a sum of exponential functions and it is well known (Lanczos (1966)) that the solutions are very difficult to find, because the functions are not orthogonal to each other. That means that in general - i.e. with more standards and more variable stimuli, the matrix of second derivatives of the likelihood function will be almost singular, causing numerical problems in the estimation process, but also giving large standard errors for the estimates, thus making the statistical tests of the model almost powerless.

5. CONCLUSION

A probabilistic model is developed that is adapted to the data collection method known as 'pick any out of n', under the condition that statistical replications of the choices are possible. The psychophysical problem of systematic errors in angle reproduction is chosen as a typical example. The basic experiment consists of a classification task, where human subjects, serving as replicates, judge for each of a series of angles (called variable angles) if they are equal or not to a given standard. The choice probability - in terms of the experiment the probability of an 'equal' judgment - as a function of the scale values of the variable angles is considered to be a product of two logistic functions. A part from the scale values two parameters - the boundaries of the acceptance region - are associated with the ideal point, which itself is not parametrized. Two other parameters - the discrimination parameters - can be interpreted as the sharpness of the boundaries.

Further, it is shown under which conditions the model is not identifiable. There are essentially two reasons: overparametrization and uncertainty about the order of the scale values, the latter being due to the single-peakedness of the choice function.

Finally it is pointed out that the model in its most general form, although identifiable, may be ill conditioned because the parameter estimation is akin to the problem of estimating sums of exponential functions. Therefore, it may turn out that only simplified versions of the model are useful in practice.

REFERENCES

[1] Andersen, E.B., Discrete statistical models with social science applications, Amsterdam, North Holland (1980).

[2] Birnbaum, A., Some latent trait models and their use in inferring an examinee's ability, in F.M. Lord and M.R. Novick, Statistical theories of mental test scores, Reading, Addison-Wesley (1968).

[3] Coombs, C.H., A theory of data, New York, Wiley (1964).

[4] Fischer, G.H., Einführung in der Theorie Psychologischer Tests, Bern, Huber (1974).

[5] Fischer, G.H., An experimental study of angular subtension, Quarterly Journal of Experimental Psychology, 21, (1969), 356-366.

[6] Green, D.M., and Swets, J.A., Signal detection theory, New York, Wiley (1966).

[7] Jastrow, J., On the judgment of angles and positions of lines, A. On the judgment of angles, American Journal of Psychology, 5 (1892), 214-217.

[8] Lanczos, C., Applied analysis, Englewood Cliffs, Prentice Hall (1956).

[9] MacRae, A.W., and Loh, H.K., Constant errors occur in matched reproduction of angles even if likely biases are eliminated, Perception and Psychophysics, 30 (1981), 341-346.

[10] Stevens, S.S., and Galanter, E.H., Ratio scales and category scales for a dozen perceptual continua, Journal of Experimental Psychology, 54 (1957), 377-411.

[11] Torgerson, W.S., Theory and methods of scaling, New York, Wiley (1958).

This Page Intentionally Left Blank

PART II

ORDER AND MEASUREMENT

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

ABOUT THE ASYMMETRIES OF SIMILARITY JUDGMENTS: AN ORDINAL POINT OF VIEW

Jean-Pierre Barthélémy
Ecole Nationale Supérieure des Télécommunications
Paris, France

We propose to revisit the notion of asymmetry of similarity judgments, with a purely ordinal point of view. We will use the majority rule, compute median orders and then compare them in order to detect (ordinal) asymmetries. An experiment with microcomputers is analyzed.

1. INTRODUCTION

The asymmetrical relationships in the context of similarity judgments play some part in attribution theory (cf. S. Duval and V.H. Duval (1983)). Indeed it follows from Tversky's paper "Features of similarity" (1977) that a strong connection exists between the asymmetry of a relation, the salience of a stimulus and the notion of prototypicality.

This paper sets up a technical reflexion about this notion of asymmetry coming from the following remark:
assume that perceived similarities between four stimuli a,b,c,d are rated in two experiments: S(a,b) (the similarity of a to b) and S(c,d) are rated in experiment 1; S(b,a) and S(d,c) in experiment 2.

A numerical asymmetry is observed when, for instance, $S(a,b) \neq S(b,a)$ (noticeably).

Assume now that the experiment 1 leads to the inequality $S(a,b) < S(c,d)$ (the perceived similarity of c to d exceeds the perceived similarity of a to b). Then from experiment 2 we can get either
(i) $S(b,a) < S(d,c)$, or
(ii) $S(d,c) \leq S(b,a)$.

In the first case the action of asymmetry remains purely local; it just involves the pairs of stimuli, without disturbing in deepth the perceived similarity itself. In the second case, that notion of similarity (viewed as a ranking operator on the pairs of stimuli) is disrupted.

Moreover, we know from traditions borrowed from scaling and measurement and from the common sense as well, that similarity analyses are often founded on similarity orders on pairs.

For all those reasons we propose, in this paper, to revisit the notion of asymmetry with a purely ordinal point of view. The first part develops points sketched in this introduction. The second part presents an experiment analyzed in the third part.

2. FROM NUMERICAL ASYMMETRY TO ORDINAL ASYMMETRY

2.1. TVERSKY'S APPROACH

Asymmetry in the context of similarity judgments has been studied by Tversky (1977) (c.f. also Tversky and Gati (1978)). In Tversky's paper asymmetries are observed with the help of classical statistics (Student t). The involved similarities are directional: they are rated on ordered pairs (a,b), with one subject a and one referent b. Tversky's hypothesis is that the direction of asymmetry is determined by the relative salience of the stimuli: the variant (subject) is more similar to the prototype (referent) than vice versa. Tversky relates several experiments supporting this hypothesis.

Moreover, he proposes a model involving those asymmetries: a and b being two stimuli, the similarity S(a,b) of a to b is expressed as a function of three arguments:

$A \cap B$ the features which are common to both a and b;
$A \setminus B$ the features that belong to a but not to b;
$B \setminus A$ the features that belong to b but not to a.

Moreover, $S(a,b) \leq S(c,d)$ whenever

$A \cap B \subset C \cap D,\ A \setminus B \supseteq C \setminus D,\ B \setminus A \supseteq D \setminus C.$

Hence Tversky's general model may be expressed as follows:
$S(a,b) = F(A \cap B, A \setminus B, B \setminus A)$, F being monotonic in the above sense. Introducing additional assumptions about the similarity ordering, Tversky obtains a representation of the general model by the contrast model:
$S^*(a,b) = \alpha f(A \cap B) - \lambda f(A \setminus B) - \mu f(B \setminus A)$. In a directional task, λ and μ may differ and consequently asymmetry may not hold.

Indeed, the obtention of that model depends upon the ordinal properties of similarities: the values S(a,b), computed from the observations, just play the role of a medium for the obtention of the complete preorder $\leqslant$ on the set of pairs of stimuli:

$$(a,b) \leqslant (c,d) \text{ iff } S(a,b) \leqslant S(c,d).$$

That preorder is represented by the model. Moreover, it contains clearly the most reliable information about the similarities, the numerical data S(a,b) being possibly noised. Nevertheless, it is in those numerical data that Tversky looks for asymmetry. If one has the desire to avoid the problem linked to the noise probably present in the data, one could propose a notion of asymmetry based only upon the preorder $\leqslant$. But this ordinal point of view is mainly supported by the deeply ordinal aspect of a type of asymmetry as emphasized previously in the introduction.

2.2. AN APPROACH BASED UPON THE AVERAGE ORDERINGS

In Tversky's paper, S(a,b) is the mean value of individual ratings. First keep that way to obtain similarities.

Consider the following situation: two groups are questionned for each pair a,b. The group G_1 works on judging the similarity of a to b, the group G_2 works on the similarity of b to a. The mean values S(a,b) and S(b,a) are computed for each group: S(a,b) for G_1, S(b,a) for G_2. Therefore, to each group corresponds a complete preorder induced by S.

We say that the two groups are in a symmetrical judgment when the two complete preorders coincide. The comparison of those preorders will give us information about the importance and the location of asymmetries. Under that name of asymmetry, we understand here, any deviation between these two orderings.

We notice some weak points at that approach; let's mention two of them. The first will lead to some care in use. The second will suggest a method suitable with our ordinal point of view.

2.3. A DRAWBACK: THE INFLUENCE OF THE DISTRIBUTION OF THE PAIRS OF STIMULI

Following again Tversky's line, let's imagine that the distribution between G_1 and G_2, of the ordered pairs of stimuli is such that: subjects in G_1 meet only the ordered pairs (a,b) where b is "more salient" than a, subjects in G_2 meeting the inverted situation.

Then we can expect that the S(a,b) (in G_1) would be greater than the S(b,a) (in G_2). But the two orderings would perhaps coincide (in that case the observed asymmetries would be only numerical but not ordinal). However, another distribution of the ordered pairs between G_1 and G_2, retaining the same numerical asymmetries, would lead possibly to distinct orderings.

In order to observe all possible ordinal asymmetries, we need to look at the 2n-1 possible distributions of our n pairs. For lack of meeting such a design, it seems desirable to deal with an experiment fitting the two following properties.

(a) A priori: the ordered pairs (a,b) where b is salient, are equitably distributed between the two groups.

(b) A posteriori: if those a priori saliences do not lead to asymmetries (according to Tversky's interpretation) one could however wish to obtain, as a way of control, an equitable distribution of the ordered pairs (a,b) so that $s(a,b) > s(b,a)$ (after having standardized the variable S into s) between the two groups.

2.4. ANOTHER DRAWBACK AND ITS REMEDY

Consider now the obtention of the two orderings. They are constructed from the mean values S(a,b). Hence we rest on the rates given by each person (individual level). Those ratings are "forgotten" at the collective level at the benefit of an order. Under the assumption that those numerical values are noised, some looseness is to be fared on the orders themselves. In this respect it seems more convenient to found the collective ordering over individual orderings. But this approach raises a difficult problem (illustrated for instance, by the famous Arrow's theorem). We propose here under the name of condorcean analysis, to use the majority rule and to compute median orders and then to compare them in order to detect (ordinal) asymmetries.

To each group is associated a linear order on the pairs builded up from the orders induced by the ratings given by each individual. That is to say that the drawback mentioned above in 2.3 remains, and some prudence will command attention.

3. THE EXPERIMENTAL DATA

3.1. THE BOOKLETS

Fourteen pairs of microcomputers have been settled (with twenty machines). They have been chosen with the help of a first investigation beside experts in microcomputing (according to those experts, the pairs do not present any drastic dissimilarity feature). Moreover, ten among them are such that a stimulus (i.e. a microcomputer) is more "dazzling", "known", "standard",... than the other. We say that such a pair admits an a priori salience. The four other pairs are said to be a priori neutral.

Each pair a,b produces two ordered pairs (a,b) and (b,a). When a,b admits an a priori salience, we say, when b is more salient than a, that (a,b) is in (-,+) situation, while (b,a) is in (+,-) situation.

The twentyeight ordered pairs so formed are divided into two booklets (one green, one blue) so that:

- each pair occurs one and only one time in each booklet;
- each booklet contains exactly five ordered pairs in (-,+) situation and five ordered pairs in (+,-) situation.

The order for introducing the pairs has been randomized in each booklet. For the persons in the two groups, the instruction is, for each ordered pair (a,b): "give a mark indicating the degree of similarity of a to b". In order to do so, the stimuli a and b are consecutively introduced with some descriptive features: constructor, country, composition of the central unit (ROM, RAM, processor), applications and languages. It is well noticing that those features correspond to a basic product (not involving any extension possibility).

3.2. GROUPS AND SUBJECTS

In order to take care of the interindividual variability we have adopted a match group procedure by controlling the professions and the levels of knowledge in informatics.

The subjects are divided into two groups. The first receive the blue booklet, the second the green booklet. The first contains 31 persons, the second 41.

3.3. NON INFLUENCE OF A PRIORI SALIENCES ON THE NUMERICAL ASYMMETRIES

For each pair a,b the following numbers have been computed:

S(a,b) and S(b,a): the mean values, into each group, of the degree of similarity of one stimulus to the other.

S+,- (a,b): the mean value of the ordered pair in situation (+,-) issued from a,b:(S+,- (a,b) = S(a,b) or S(b,a)); S-,+ (a,b) the value for the ordered pair in situation (-,+).

All those variables are standardized into s; s+,-; s-,+.

The examination of the values of S (c.f. appendix) indicates that, among the ten pairs with an a priori salience, the difference S-,+ (a\b) - S+,- (a\b) is positive only on four pairs. This remark may not be nicely interpretated (the averages of the two groups being different). Otherwise, using the standardized variables, this tendency reverses.

For each of the 72 booklets, the average for the five ordered pairs in (-,+) situation has been computed as well as the average for the five pairs in (+,-) situation. So two variables $T_i(-,+)$ and $T_i(+,-)$ $(1 \leq i \leq 72)$ have been obtained. Their mean values are:

$$m\ (-,+) = \frac{1}{72} \sum_{i=1}^{72} T_i\ (-,+) = 10{,}730$$

$$m\ (+,-) = \frac{1}{72} \sum_{i=1}^{72} T_i\ (+,-) = 10{,}647.$$

This leads to (Student test) $t = 0{,}186$; $p = .67$. So m (-,+) does not differ significantly from m (+,-).

It is worth noticing that this result does not contradict Tversky's interpretation. Indeed, the roles of subject and of referent are not very clear here. Although it is asked to the persons to compare a to b, they discover (and "learn") at first a, then b (if the pages of the booklets are regularly read from the top to the bottom,...).

3.4. THE DISTRIBUTION OF THE ORDERED PAIRS AND THE MAXIMAL NUMERICAL ASYMMETRY

Among the fourteen ordered pairs (a,b) in each booklet, seven are so that $s(a,b) > s(b,a)$. A posteriori the ordered pairs are, concerning the direction of the asymmetry, fairly distributed and the demand (b) of 2.3 is satisfied.

Let's compute, for each booklet the mean value of the seven $S(a,b)$ so that $s(a,b) > s(b,a)$. We obtain so the variable U_i. The variable V_i is defined from the seven $S(a,b)$ so that $s(a,b) < s(b,a)$. It becomes $1/72 \sum_{i=1}^{72} U_i = 11,242$; $1/72 \sum_{i=1}^{72} V_i = 10,523$; $t = 1,769$; $p = .038$. This grouping corresponds to the "maximal numerical asymmetry" (remember that one has to take the value of p with great prudence, keeping in mind that from the a posteriori selection of U and V, p is only a lower bound).

4. CONDORCEAN ANALYSIS

4.1. FROM CONDORCET TO THE MEDIAN PROCEDURE

Consider a set $G = \{1,\dots,n\}$ of individuals and a set X of objects. Assume that each individual i carries out a paired comparison. That comparison is typified by a binary relation R_i on X. In order to put flesh on the bone, we can interpret xR_iy as "y is preferred to x by the individual i". Considering all the R_i, we obtain the profile $\pi = (R_1,\dots,R_n)$. We denote by $\pi(x,y)$ the number of individuals preferring y to x: $\pi(x,y) = |\{i \mid xR_iy\}|$. When $\pi(x,y) \geq P/2$, y is said to be majoritary preferred to x. Therefore we obtain the so-called Condorcet relation of the profile: xRy iff y is majoritary preferred to x.

Formally we can write:

$$R = \bigcup_{\substack{G' \subseteq G \\ |G'| = [\frac{n+1}{2}]}} \ \bigcap_{i \in G'} R_i ,$$

the intersection and the union of binary relations being defined as usual. Otherwise that Condorcet relation looks like a median (c.f. Barthélémy and Monjardet, (1981)).

In our case, we try to summarize the similarity orders, given by each individual in one group, into a linear order. At first, we notice that when each R_i is complete, then R is complete too. Nevertheless R will not need to be a transitive relation even if each R_i is so. That phenomenon is known as "Effet Condorcet" (c.f. Guilbaud (1952)) or "paradox of voting". In order to reform it, it is classical to proceed as follows (when each R_i is complete): by inverting the direction of some arcs (i.e. by transforming some xRy into yRx), we are able to get, from R, a transitive relation $\bar{R}$.

Among all possible $\bar{R}$, we choose the relation minimizing the sum

$$\sum_{(x,y)\in C} (\pi(x,y) - \pi(y,x)),$$

C denoting the set of inverted arcs. It is then sufficient to break the ties (without creating any cycle) in order to obtain a linear order, the so-called median order of the profile. Properties and justifications of the use of median orders may be found in Barthélémy and Monjardet (1981).

4.2. CONDORCEAN ANALYSIS OF TWO GROUPS

Consider now two groups G_1 and G_2 of respectively n_1 and n_2 persons. At each individual i in G_k (k=1,2) is associated the binary relation on X: the "opinion" of i on the elements of X. Thus we obtain two profiles π^1 and π^2.

The condorcean analysis compares a median order associated to π^1 to a median order associated to π^2. Since a profile may admit several median orders, these two are chosen with a maximal correlation (in Kendall's sense). This comparison may involve the following practices:

(a) qualitatively: in order to locate the (ordinal) asymmetries, we propose to:

- detect the sequences (of consecutive objects) common to the two median orders and maximal;
- determine the subsets $Q_1,Q_2,\ldots,Q_p$ of X, called ordered blocks, so that for each of the two median orders $Q_i < Q_{i+1}$ (each element in Q_i is bounded by each element of Q_{i+1}), the Q_i being minimal;
- determine the intersection partial order of the two median orders and examine the incomparable elements in that intersection.

(b) quantitatively: beyond the computation of the Kendall's τ, we propose three coefficients (more or less inspirated by Gordon (1979a) and (1979b)). All those coefficients will provide an evaluation of the magnitude of the (ordinal) asymmetry. They take values between 0 and 1 and are defined such as:

$\alpha/n-1$, α being the sum of the lengths of the maximal common sequences;

$p-1/n-1$, p being the number of minimal ordered blocks;

$\beta/n-1 = n-\gamma/n-1$, β being the maximal length of a maximal chain in the intersection order and γ denoting the minimal number of elements to remove in order to make the two median orders coincide.

4.3. A REMARK ABOUT THE EXPERIMENTAL DESIGN

It is worth noticing that the essential nature of the majority rule forbids any merging - both explicit or implicit - of the two groups into one.

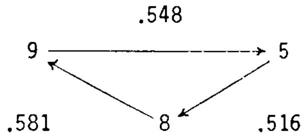
4.4. CONDORCEAN ANALYSIS OF THE SIMILARITY ORDERS: QUALITATIVE STUDY

Let 1,2,...,14, denote the fourteen pairs of microcomputers. The Condorcet relation associated to the blue booklet (group G_1) may be pictured as follows:

$$\begin{tikzcd}
3 \arrow[r] & 8 \arrow[r] & 9 \arrow[r] & 5 \arrow[r] \arrow[ll, bend left] & 2 \arrow[r] & 7 \arrow[r] & 14 \arrow[r] & 10 \arrow[r] & 13 \arrow[r] \arrow[d, shift left] & 4 \arrow[r] & 12 \arrow[r] & 1 \\
& & & & & & & & 11 \arrow[u, shift left] \arrow[d, shift left] & & & \\
& & & & & & & & 6 \arrow[u, shift left] & & &
\end{tikzcd}$$

(The arrows are composed by transitivity excepted any mention of the contrary, f.i. 8   1   5 prohibit the existence of 8   5).

That relation admits just one cycle



and the median orders are extracted, with a tie breaking, from the following complete preorder.

$$\begin{tikzcd}
3 \arrow[r] & 8 \arrow[r] & 9 \arrow[r] & 5 \arrow[r] & 2 \arrow[r] & 7 \arrow[r] & 14 \arrow[r] & 10 \arrow[r] & 13 \arrow[r] \arrow[d, shift left] & 4 \arrow[r] & 12 \arrow[r] & 1 \\
& & & & & & & & 11 \arrow[u, shift left] \arrow[d, shift left] & & & \\
& & & & & & & & 6 \arrow[u, shift left] & & &
\end{tikzcd}$$

From the green booklets we get the following Condorcet relation:

$$\begin{tikzcd}
5 \arrow[r] & 3 \arrow[r] & 8 \arrow[r] & 9 \arrow[r] & 2 \arrow[r] & 7 \arrow[r] & 14 \arrow[r] & 10 \arrow[r] & 13 \arrow[r] & 6 \arrow[r] \arrow[d, shift left] & 1 \arrow[r] & 4 \arrow[r] & 12 \\
& & & & & & & & & 11 \arrow[u, shift left] & & &
\end{tikzcd}$$

which admits no cycle.

The two following median orders are in maximal correlation.

Blue $3 < 8 < 9 < 5 < 2 < 7 < 14 < 10 < 13 < 11 < 6 < 4 < 12 < 1$

Green: $5 < 3 < 8 < 9 < 2 < 7 < 14 < 10 < 13 < 11 < 6 < 1 < 4 < 12$

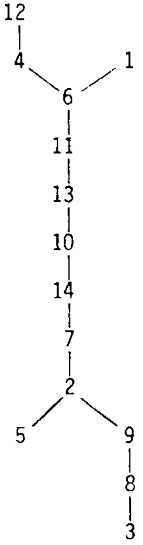
The maximal common sequences are:

$$3 < 8 < 9 \;,\; 2 < 7 < 14 < 10 < 13 < 11 < 6 \;,\; 4 < 12$$

They represent pieces of symmetrical judgments between the two groups. The middle sequence points out a complete agreement (symmetry) for seven pairs located at the middle of the scales. This sequence is dominated by a top parcel {4,12,1} and dominates a tail parcel {3,8,9,5}. In those two parcels asymmetries occur.

The decomposition of the two orders into ordered blocks corroborates those remarks. It points the top parcel O_9 = {12,4,1}, the tail parcel {3,8,9,5}, the middle sequence being broken up into seven singletons.

The Hasse diagram of the intersection of the two median orders is pictured. It corroborates the previous remarks and points out asymmetries involving the following pairs: 12 and 1, 4 and 1 for the stimuli perceived as the most similar; 5 and 9, 5 and 8, 5 and 3 for the stimuli perceived as the less similar. The asymmetry at the top is diverting enough. It concerns the three pairs involving the same three microcomputers: 1 = TRS 80 mod 2/ GENIE 3, 4 = Z 89 / GENIE 3, 12 = Z 89 / TRS 80 mod 2.



4.5. CONDORCEAN ANALYSIS: QUANTITATIVE STUDY

The values of the coefficients we have proposed are:

$$\tau = .89, \quad \alpha/n-1 = .69, \quad \frac{p-1}{n-1} = .62, \quad \beta/n-1 = .85.$$

They indicate a large enough symmetry, each of them involving its proper information.

4.6. ANALYSIS OF THE AVERAGE ORDERS

The mean values of the rates given to each pair of microcomputers, induce on each booklet the following complete preorders.

Blue: $3 < 8 < 5 < 9 < 2 < 7 < 14 < 10 < 11 < \substack{13 \\ 6} < 12 \lesssim 4 < 1$

Green: $5 < 3 < 8 < 9 < 7 < 2 < 14 < 10 < 13 < 11 < 6 < 4 < 1 < 12$

($\lesssim$ means a 1% difference).

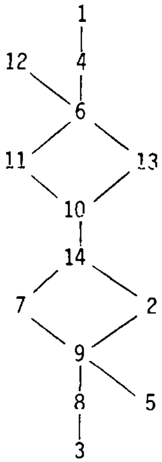
Comparing to the median orders we find a coincidence excepted for: two inversions 2, 7 and 1, 4 for the green booklets; two inversions 9, 5 and 4, 12 for the blue booklets.

The asymmetries are better marked than in the case of median orders.

The maximal common sequences are:
$3 < 8$, $14 < 10$, $11 < 6$ and $4 < 1$. The decomposition into ordered blocks leads to eight Q_i.

$Q_1 = \{3, 5, 8\}$, $Q_2 = \{9\}$, $Q_3 = \{7, 2\}$, $Q_4 = \{14\}$, $Q_5 = \{10\}$,
$Q_6 = \{13, 11\}$, $Q_7 = \{6\}$, $Q_8 = \{4, 1, 12\}$.

The Hasse diagram of the intersection is pictured. That partial order is contained into the intersection of the two median orders.



Kendall τ remains high: $\tau = .87$

The other coefficients decrease: $\alpha/n-1 = .31$, $p-1/n-1 = .54$, $\beta/n-1 = .69$.

4.7. AS A (PROVISIONAL) CONCLUSION

From the previous study the following points emerge.

(i) In a context not so propitious as those surrounding Tversky's experiment:

- the notion of salience is not so clear in the field of microcomputers,
- the notions of subject and referent are ambiguous; asymmetries keep on to be observed but in a weaker sense, either numerical with an optimal grouping (which does not coincide with the a priori saliences) or ordinal (indeed, that weakness confirms, in some sense, Tversky's interpretation).

(ii) The ordinal analysis emphasizes the fact that the asymmetries are related to the position of the pairs of stimuli in the similarity scale. They are observed essentially (entirely in the case of the median orders) at the head and the tail of that scale.

(iii) Of course, the observed asymmetries depend upon the observation methods. They are lower for median orders than for average orders. Now, the former are much more numerical than the latter. Is this any connection between the degree of "numericality" of the orders and the possibility of observing asymmetries?

REFERENCES

[1] Barthélémy, J.P., and Monjardet, B, The median procedure in cluster analysis and social choice theory, Math. Soc. Sci., 1 (1981), 235-267.

[2] Duval, S, and Duval, V.H., Consistency and cognition, Lawrence Erlbaun ass., Hillsdale, London (1983).

[3] Gordon, A.G., A measure of the agreement between rankings, Biometrika, 66, 1 (1979a), 7-15.

[4] Gordon, A.G., Another measure of the agreement between rankings, Biometrika, 66, 2 (1979b), 327-332.

[5] Guilbaud, G.Th., Les théories de l'intérêt général et le problème logique de l'agrégation, Economie Appliquée, 5(4), (1952). Reprinted in Elements de la théorie des jeux, Dunod, Paris (1968), English translation in, Readings in Mathematical social sciences, Science Research Associates, Chicago (1966), 262-307.

[6] Kendall, M.G., Rank correlation methods, Hafner, New York, 3rd edition (1962).

[7] Tversky, A., Features of similarity, Psychological Review, 84, 4 (1977) 327-352.

[8] Tversky, A, and Gati, I, Studies of similarity, in E. Rosh and B. Lloyd, eds., On the nature and the principle of formation of categories, Erlbaun, Hillsdale and London.(1979).

APPENDIX

The fourteen pairs

1	TRS 80 Mod. 2	/ GENIE 3
2	SIL'Z 3	/ APPLE II Plus
3	H P 85 A	/ OSBORNE 01
4	Z 89	/ GENIE 3
5	TI 99/4A	/ ZX 81
6	ABC 24	/ TRS 80 Mod. 2
7	XEROX 820	/ APPLE III
8	TX 25	/ GOUPIL 2
9	VIC 20	/ DAI
10	M 20	/ ITT 3030
11	XEROX 820	/ MICRAL 8021
12	Z 89	/ TRS 80 Mod. 2
13	APPLE III	/ MICRAL 8021
14	DAI	/ VICTOR LAMBDA

The writing order of each pair and the order between the pairs are those of the blue booklets.

The a priori salience

In the blue booklets the a priori saliences were so constituted.

(+,-): 1, 3, 4, 9, 13
(-,+): 2, 6, 7, 8, 14.

Statistical description of the booklets

Pair	mean similarity		standard deviation		standardized similarity	
a/b	Blue	Green	Blue	Green	Blue	Green
1	13.32	14.17	2.428	2.594	1.35	1.22
2	9.37	10.80	3.440	4.080	-0.54	-0.17
3	6.45	7.96	4.211	4.750	-1.93	-1.35
4	12,52	13.73	3.202	3.548	0.96	1.04
5	8.20	7.07	3.987	3.703	-1.10	-1.71
6	12,45	13.39	3.416	3.341	0.93	0.90
7	9.71	9.79	3.786	3.918	-0.38	-0.59
8	7.97	8.16	4.701	3.606	-1.21	-1.26
9	8.32	9.22	3.237	3.524	-1.04	-0.83
10	11.06	11.51	4.165	3.762	0.27	0.12
11	12.21	13.38	4.288	3.391	0.82	0.89
12	12.48	14.95	3.048	3.107	0.94	1.54
13	12.45	11.71	3.582	3.983	0.93	0.20
14	10.45	11.22	3.867	3.745	-0.02	0.00
	10.497	11.220	2.097	2.420	0	0

Optimal grouping

Ordered pairs (a,b) so that s(a,b) > s(b,a): 1, 5, 6, 7, 8, 10, 13 (for the conventional writing order of each pair).

Namely	A priori salience	Agreement with a priori salience
(TRS 80 Mod 2, GENIE 3)	(+,-)	NO
(APPLE II$^+$, SIL'23)	(+,-)	NO
(OSBORNE 01, HP 85A)	(-,+)	YES
(GENIE 3, Z 89)	(-,+)	YES
(TI 99/4A, ZX 81)	neutral	
(ABC 24, TRS 80 Mod 2)	(-,+)	YES
(XEROX 820, APPLE III)	(-,+)	YES
(TX 25, GOUPIL 2)	(-,+)	YES
(DAI, VIC 20)	(-,+)	YES
(M20, ITT 3030)	neutral	
(MICRAL 8021, XEROX 820)	neutral	
(TRS 80 Mod 2, Z 89)	neutral	
(APPLE III, MICRAL 8021)	(+,-)	NO
(VICTOR LAMBDA, DAI)	(+,-)	NO

The author wishes to thank Hervé Abdi, Caroline Devarenne, Yves Le Foll and Jawad Mourtada for their help.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

THE AXIOMATIZATION OF ADDITIVE DIFFERENCE MODELS FOR PREFERENCE JUDGMENTS

M.A. Croon

Vakgroep Statistiek en Methoden & Technieken
van Onderzoek
Subfaculteit Psychologie
Katholieke Hogeschool Tilburg

Following the approach taken by Fishburn (1980), the algebraic axiomatization of an additive difference model for pairwise choices between multi-attribute stimuli is discussed.

1. INTRODUCTION

Tversky (1969) showed that, under specified experimental conditions, consistent and predictable intransitivities in the pairwise choices between multi-attribute stimuli can be demonstrated. Furthermore, he discussed two models that can account for these results. The first model which imposes a just noticeable difference structure or semiorder on a lexicographic ordering, was used by him in the construction of the alternatives for his experimental investigations. A similar model was studied by Ng (1977). The second model, called the additive difference model, assumes that the choice between pairs of multi-attribute stimuli is based on comparisons of component-wise differences between the alternatives. Let $A=A_1 x \dots x A_n$ be a set of multi-attribute stimuli with elements of the form $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$ where x_i represents the value of stimulus x on attribute i. Let $\gtrsim_0$ be the empirical preference relation on A such that $x \gtrsim_0 y$ if and only if x is at least as much preferred as y. The preference structure $\langle A,\gtrsim_0\rangle$ satisfies the additive difference model if there exist real-valued functions $u_1,\dots,u_n$ and increasing functions $\phi_1,\dots,\phi_n$ defined on some real intervals such that $x \gtrsim_0 y$ if and only if:

$$\sum_{i=1}^{n} \phi_i[u_i(x_i)-u_i(y_i)] \geq 0$$

where $\phi_i(-\delta)=-\phi_i(\delta)$ for all i.

Although Tversky (1969) discussed some properties of this additive difference model, he did not provide a complete axiomatization of it although he mentioned that a set of ordinal axioms for a symmetric additive difference model of similarity (rather than preference) judgments was given by Beals, Krantz and Tversky (1968).

As a model of preference judgments, the additive difference model received little attention. In his dissertation, van Acker (1977) reviews the literature on intransitive choice and mentions the additive difference model and some generalizations of it, without giving a thorough discussion of their axiomatization. Recently Fishburn (1980) provided an axiomatization of the so-called lexicographic additive difference model in which preference judgments are globally governed by a lexicographic rule which may locally (e.g. when the two stimuli are quite similar) be superseded by an additive difference rule. Moreover, the pure additive difference model can be considered as a limit case of this more general model. In his paper, Fishburn restricts himself to the case in which the stimuli are composed of two attributes. Furthermore, his axiomatization is based on rather strong topological assumptions about the component sets.

In the present paper, Fishburn's approach to the axiomatization of the pure additive difference model is pursued. In the first place, an attempt was made to replace the topological assumptions about the component sets by some more algebraic assumptions so as to provide an axiomatization of the pure additive difference model which is more in line with the approach of Krantz et al. (1971). In the second place, this paper shows how Fishburn's approach can be generalized to the case when the stimuli are composed of more than two components. Due to space limitations, we only give full proofs of lemmas and theorems for the case of two components. The proofs for the case of three component sets are briefly sketched. The general multi-attribute additive difference model and some generalizations of it will be discussed in a subsequent paper.

2. ADDITIVE DIFFERENCE STRUCTURES ON TWO COMPONENTS SETS

In the sequel we assume that A and X are two non-empty sets and that $\geq_0$ is a binary relation on AxX. Elements of A will be denoted by small letters, either primed or unprimed, from the first half of the alphabet; similarly, small letters from the second part of the alphabet will be used to denote elements of X. Finally, elements of AxX will be denoted by ordered pairs of small letters such as ax, by, cu, etc.

The structure $\langle AxX, \geq_o \rangle$ is said to be an additive difference structure on two component sets if the binary relation $\geq_o$ satisfies the following axioms:

1. Axiom 1

The binary relation $\geq_o$ is reflexive and connected:

(1) $ax \geq_o ax$ for all $ax \in AxX$

(2) $ax \geq_o by$ or $by \geq_o ax$ for all $ax, by \in AxX$.

Note that we do not assume that $\geq_o$ is transitive. Hence, $\geq_o$ is not a weak order relation.

2. Axiom 2

The relation $\geq_o$ satisfies the following triple cancellation (TC) condition:

$$\left.\begin{array}{l} aw \geq_o bx \\ by \geq_o az \\ cz \geq_o dy \end{array}\right\} \Rightarrow cw \geq_o dx$$

3. Axiom 3

The relation $\geq_o$ satisfies the following quintuple cancellation conditions:

(1) QCA

if $\left\{\begin{array}{l} au \geq_o bv \\ bx \geq_o cy \end{array}\right.$ and $\left\{\begin{array}{l} b'v \geq_o a'u \\ c'y \geq_o b'x \end{array}\right.$

then $cz \geq_o aw$ implies $c'z \geq_o a'w$

(2) QCX

if $\left\{\begin{array}{l} au \geq_o bv \\ cv \geq_o dw \end{array}\right.$ and $\left\{\begin{array}{l} bv' \geq_o au' \\ dw' \geq_o cv' \end{array}\right.$

then $fw \geq_o gu$ implies $fw' \geq_o gu'$.

In order to formulate the next axioms we define two new relations $=_o$ and $>_o$.

Definition 1

$ax =_o by$ iff $ax \geq_o by$ and $by \geq_o ax$

One easily proves that $=_o$ is reflexive and symmetric. However, $=_o$ is not necessarily transitive: hence it is not an equivalence relation.

Definition 2

$ax >_o by$ iff $ax \geq_o by$ and not $by \geq_o ax$

The relation $>_o$ is irreflexive.

We now formulate the solvability conditions.

4. Solvability conditions

(1) Unrestricted solvability for two unknowns:

a. USA: given $x,y \in X$, there exist $a,b \in A$ such that $ax =_o by$.

b. USX: given $a,b \in A$, there exist $x,y \in X$ such that $ax =_o by$.

(2) Restricted solvability for one unknown:

a. RSA: if $ax \geq_o by$ and $by \geq_o cx$, there exists $d \in A$ such that $dx =_o by$.

b. RSX: if $ax \geq_o by$ and $by \geq_o az$, there exists $u \in X$ such that $au =_o by$.

5. Archimedean axiom

In order to formulate this axiom, we need some further definitions.

A sequence $(a_1,a_2,\ldots,a_n,\ldots)$ of elements of A is called a standard sequence for component set A if there exist $y, z \in X$ such that for all n:
$a_n y =_o a_{n-1} z$.

A standard sequence is strictly bounded if there exist $b_1,b_2 \in A$ such that for all n there exist $u_n,v_n,w_n,x_n \in X$ such that:

(1) $b_1 u_n >_o b_2 v_n$ and $a_1 v_n >_o a_n u_n$

(2) $a_n w_n >_o a_1 x_n$ and $b_1 x_n >_o b_2 w_n$.

Similar definitions hold for the second component set X. Then the archimedean axiom can be stated in the following way: every strictly bounded standard sequence is finite.

For further use we prove the following

LEMMA 1

$\geq_o$ satisfies independence:

(1) $aw \geq_o bw$ implies $ax \geq_o bx$

(2) $aw \geq_o ax$ implies $bw \geq_o bx$

Proof:

(1) Suppose aw $\geq_o$ bw.
Also ax $=_o$ ax and aw $=_o$ aw.
Then by TC

$$\left.\begin{array}{l} ax =_o ax \\ aw =_o aw \\ aw \geq_o bw \end{array}\right\} \text{imply } ax \geq_o bx$$

(2) Suppose aw $\geq_o$ ax.
Then by TC

$$\left.\begin{array}{l} bw =_o bw \\ aw =_o aw \\ aw \geq_o ax \end{array}\right\} \text{imply } bw \geq_o bx$$

We now prove that the axioms 1 to 5 are sufficient for an additive difference representation of $\geq_o$. Given that the binary relation $\geq_o$ satisfies these conditions, there exist a function f on A, a function g on X and a function ϕ such that

$$ax \geq_o by \text{ iff } \phi[f(a)-f(b)] + g(x)-g(y) \geq 0.$$

The first step in this proof consists of defining relations R_A on AxA and R_X on XxX in the following way:

Definition 3

For arbitrary a,b,c,d $\in$ A, (a,b) R_A (c,d) holds if there exist x,y $\in$ X such that ax $\geq_o$ by and dy $\geq_o$ cx.
For arbitrary w,x,y,z $\in$ X, (w,x) R_X (y,z) holds if there exist a,b $\in$ A such that aw $\geq_o$ bx and bz $\geq_o$ ay.

We can now prove that $\langle AxA,R_A\rangle$ and $\langle XxX,R_X\rangle$ both satisfy the axioms of an algebraic-difference structure as stated by Krantz et al. (1971, 150-154). We give this proof only for $\langle AxA,R_A\rangle$.

1. R_A is a weak order.

a. R_A is reflexive.
Given a,b $\in$ A, by USX there exist x,y such that ax $=_o$ by. Then also by $=_o$ ax. So (a,b)R_A(a,b).

b. R_A is connected.
Given a,b $\in$ A, by USX there exist x,y such that ax $=_o$ by. Then either (1) cx $\geq_o$ dy or (2) dy $\geq_o$ cx.

In case (1): $cx \geq_o dy$ and $by =_o ax$ imply $(c,d)R_A(a,b)$.

In case (2): $ax =_o by$ and $dy \geq_o cx$ imply $(a,b)R_A(c,d)$.

c. R_A is transitive.

Suppose $(a,b)R_A(c,d)$ and $(c,d)R_A(e,f)$.

Now:

$(a,b)R_A(c,d)$ implies $av \geq_o bw$ and $dw \geq_o cv$ for some $v,w \in X$ and

$(c,d)R_A(e,f)$ implies $cx \geq_o dy$ and $fy \geq_o ex$ for some $x,y \in X$.

By TC:
$$\left.\begin{array}{l} dw \geq_o cv \\ cx \geq_o dy \\ fy \geq_o ex \end{array}\right\} \text{imply } fw \geq_o ev$$

Then: $av \geq_o bw$ and $fw \geq_o ev$.

Hence $(a,b)R_A(e,f)$.

Definition 4

$(a,b)I_A(c,d)$ iff $(a,b)R_A(c,d)$ and $(c,d)R_A(a,b)$.

Note: Since R_A is a weak order, it follows that I_A is an equivalence relation.

LEMMA 2

$(a,b)I_A(c,d)$ iff $au =_o bv$ and $cu =_o dv$ for some $u,v \in X$.

Proof:

(1) $(a,b)I_A(c,d)$ implies $(a,b)R_A(c,d)$ and $(c,d)R_A(a,b)$.

Now

$(a,b)R_A(c,d)$ implies $au \geq_o bv$ and $dv \geq_o cu$ for some $u,v \in X$ and

$(c,d)R_A(a,b)$ implies $cx \geq_o dy$ and $by \geq_o ax$ for some $x,y \in X$.

By TC:
$$\left.\begin{array}{l} au \geq_o bv \\ by \geq_o ax \\ cx \geq_o dy \end{array}\right\} \text{imply } cu \geq_o dv$$

Hence: $cu =_o dv$ since also $dv \geq_o cu$.

Furthermore, also by TC:
$$\left.\begin{array}{l} dv \geq_o cu \\ cx \geq_o dy \\ by \geq_o ax \end{array}\right\} \text{imply } bv \geq_o au$$

Hence: $au =_o bv$ since also $au \geq_o bv$.

(2) From $au =_o bv$ and $cu =_o dv$, it follows that $(a,b)I_A(c,d)$.

2. If $(a,b)R_A(c,d)$, then $(d,c)R_A(b,a)$.

Proof:

$(a,b)R_A(c,d)$ implies $ax \geq_o by$ and $dy \geq_o cx$ for some $x,y \in X$.

Then also $dy \geq_o cx$ and $ax \geq_o by$

or $(d,c)R_A(b,a)$.

3. If $(a,b)R_A(a',b')$ and $(b,c)R_A(b',c')$, then $(a,c)R_A(a',c')$.

Proof:

By hypothesis: $au \geq_o bv$ and $b'v \geq_o a'u$ for some $u,v \in X$

and: $bx \geq_o cy$ and $c'y \geq_o b'x$ for some $x,y \in X$.

By USX: $aw =_o cz$ for some $w,z \in X$.

By QCA: If $\left\{\begin{array}{l} au \geq_o bv \\ bx \geq_o cy \end{array}\right.$ and $\left\{\begin{array}{l} b'v \geq_o a'u \\ c'y \geq_o b'x \end{array}\right.$

then $cz =_o aw$ implies $c'z \geq_o a'w$.

Hence: $aw =_o cz$ and $c'z \geq_o a'w$,

or: $(a,c)R_A(a',c')$.

4. If $(a,b)R_A(c,d)R_A(a,a)$, then:

(1) there exists $d' \in A$ such that $(a,d')I_A(c,d)$

(2) there exists $d'' \in A$ such that $(d'',b)I_A(c,d)$

Proof:

By USX: $cu =_o dv$ for some $u,v \in X$.

By hypothesis: $ax \geq_o by$ and $dy \geq_o cx$ for some $x,y \in X$.

Then by TC: $\left.\begin{array}{l} cu =_o dv \\ dy \geq_o cx \\ ax \geq_o by \end{array}\right\}$ imply $au \geq_o bv$

Also: $(c,d)R_A(a,a)$ implies $cp \geq_o dq$ and $aq \geq_o ap$ for some $p,q \in X$.

(1) by TC: $\left.\begin{array}{l} aq \geq ap \\ cp \geq dq \\ dv =_o cu \end{array}\right\}$ imply $av \geq_o au$

Hence we have $av \geq_o au$ and $au \geq_o bv$.

By RSA there exists $d' \in A$ such that $au =_o d'v$.

By Lemma 2: $au =_o d'v$ and $cu = dv$ implies $(a,d')I_A(c,d)$.

(2) By hypothesis: $cp \geq_o dq$ and $aq \geq_o ap$.

By Lemma 1 (independence): $bq \geq_o bp$.

By TC: $\left.\begin{array}{l} bq \geq_o bp \\ cp \geq_o dq \\ du =_o cu \end{array}\right\}$ imply $bv \geq_o bu$

Hence we have $au \geq_o bv$ and $bv \geq_o bu$.

By RSA there exists $d'' \in A$ such that $d''u =_o bv$.

Hence $d''u =_\circ bv$ and $cu =_\circ dv$
or $(d'',b)I_A(c,d)$.

5. Archimedean property.

Let $(a_1,a_2,\ldots,a_i,\ldots)$ be a sequence of elements such that:

(1) $(a_{i+1},a_i)I_A(a_2,a_1)$ for all a_i,a_{i+1}

(2) not $(a_2,a_1)I_A(a_1,a_1)$

(3) there exist d',d'' such that $(d',d'')P_A(a_i,a_1)P_A(d'',d')$, P_A being the asymmetric part of R_A.

Then we have to prove that the sequence is finite.
We first prove that there exist $u,v \in X$ such that $a_{i+1}u =_\circ a_iv$ for all a_i,a_{i+1}.
Now: $(a_{i+1},a_i)I_A(a_2,a_1)$ implies $a_{i+1}u =_\circ a_iv$
and $a_2u =_\circ a_1v$ for some $u,v \in X$.
Also: $(a_{i+2},a_{i+1})I_A(a_2,a_1)$ implies $a_{i+2}x =_\circ a_{i+1}y$
and $a_2x =_\circ a_1y$ for some $x,y \in X$.
By TC:

$$\left.\begin{array}{l} a_{i+2}x =_\circ a_{i+1}y \\ a_1y =_\circ a_2x \\ a_2u =_\circ a_1v \end{array}\right\} \text{ imply } a_{i+2}u =_\circ a_{i+1}v.$$

Hence there exist $u,v \in X$ such that $a_{i+1}u =_\circ a_iv$ for all a_i,a_{i+1} in the sequence.
Hence, $(a_1,a_2,\ldots,a_n,\ldots)$ is a standard sequence in the sense of Axiom 5. The proof that this sequence is strictly bounded, follows from the existence of the elements d' and d'' with the stated properties. Hence, by Axiom 5 this sequence is finite, which proves that the archimedean axiom is satisfied in $\langle A\text{x}A,R_A\rangle$.

We have proved that $\langle A\text{x}A,R_A\rangle$ satisfies the axioms of an algebraic difference structure. Hence, there exists a function $f\colon A \to Re$ such that $(a,b)R_A(c,d)$ iff $f(a)-f(b) \geq f(c)-f(d)$.
In a similar way we can prove that $\langle X\text{x}X,R_X\rangle$ also satisfies the axioms of an algebraic difference structure. Hence, there exists a function $g\colon X \to Re$ such that $(u,v)R_X(w,x)$ iff $g(u)-g(v) \geq g(w)-g(x)$.
Now let $\Omega_A = \{r : r=f(a)-f(b) \text{ for some } a,b \in A\}$
and
$\Omega_X = \{s : s=g(u)-g(v) \text{ for some } u,v \in X\}$.
We define $\phi\colon \Omega_A \to \Omega_X$ as follows:

$$\phi(r) = s \text{ if } r = f(a)-f(b)$$
$$s = g(u)-g(v)$$
$$\text{and } av =_o bu.$$

Then we can prove the following lemmas.

LEMMA 3 ϕ is well-defined.

Let $r = f(a)-f(b) = f(c)-f(d)$ and suppose $av =_o bu$ and $cx =_o dw$. We have to prove: $g(u)-g(v) = g(w)-g(x)$.

Now $f(a)-f(b) = f(c)-f(d)$ implies $(a,b)I_A(c,d)$. So there exists $y,z \in X$ such that $ay =_o bz$ and $cy =_o dz$.

Now: $ay =_o bz$ and $bu =_o av$ imply $(z,y)I_X(u,v)$

and: $cy =_o dz$ and $dw =_o cx$ imply $(z,y)I_X(w,x)$.

Hence $(u,v)I_X(w,x)$ or $g(u)-g(v) = g(w)-g(x)$.

LEMMA 4 $\phi(-r) = -\phi(r)$.

Let $r = f(a)-f(b)$

$s = g(u)-g(v)$

and $av =_o bu$.

Hence $\phi(r) = s$.

Then also $bu =_o av$

$-r = f(b)-f(a)$

$-s = g(v)-g(u)$.

So $-s = \phi(-r)$ or $\phi(-r) = -\phi(r)$.

LEMMA 5 ϕ is increasing.

Let $r = f(a)-f(b)$, $r' = f(c)-f(d)$

$s = g(u)-g(v)$, $s' = g(w)-g(x)$

and $av =_o bu$, $cx =_o dw$.

Hence $\phi(r) = s$ and $\phi(r') = s'$.

Suppose $r \geq r'$.

Then $f(a)-f(b) \geq f(c)-f(d)$

or $(a,b)R_A(c,d)$.

Hence $ay \geq_o bz$ and $dz \geq_o cy$ for some $y,z \in X$.

By TC:
$$\left.\begin{array}{l} ay \geq_o bz \\ dz \geq_o cy \\ cx =_o dw \end{array}\right\} \text{ imply } ax \geq_o bw$$

Now $aw \geq_o bx$ and $bv =_o au$ imply $(w,x)R_X(u,v)$

or $g(w)-g(x) \geq g(u)-g(v)$

or $s \geq s'$
or $\phi(r) \geq \phi(r')$.

We can now prove the following Representation Theorem:

$$ax \geq_o by \text{ iff } \phi[f(a)-f(b)] + g(x)-g(y) \geq 0$$

Proof:

Part I.

Suppose $ax \geq_o by$.
By USA: $cx =_o dy$ for some, $c,d \in A$.
Then $\phi[f(c)-f(d)] = g(y)-g(x)$.
Furthermore: $ax \geq_o by$ and $dy =_o cx$ imply $(a,b)R_A(c,d)$
or: $f(a)-f(b) \geq f(c)-f(d)$.
Since ϕ is increasing: $\phi[f(a)-f(b)] \geq \phi[f(c)-f(d)]$.
Hence $\phi[f(a)-f(b)] \geq g(y)-g(x)$
or $\phi[f(a)-f(b)] + g(x)-g(y) \geq 0$.

Part II.

Suppose $\phi[f(a)-f(b)] + g(x)-g(y) \geq 0$.
By USX: $au =_o bv$ for some $u,v \in X$.
Then $\phi[f(a)-f(b)] = g(v)-g(u)$.
By USA: $cx =_o dy$ for some $c,d \in A$.
Then $\phi[f(c)-f(d)] = g(y)-g(x)$.
By assumption: $\phi[f(a)-f(b)] \geq g(y)-g(x)$.
Hence $g(v)-g(u) \geq g(y)-g(x)$
or $(v,u)R_X(y,x)$.
Hence there exist a',b': $a'v \geq_o b'u$ and $b'x \geq_o a'y$.

$$\text{By TC: } \left.\begin{array}{l} au =_o bv \\ a'v \geq_o b'u \\ b'x \geq_o a'y \end{array}\right\} \text{ imply } ax \geq_o by$$

3. ADDITIVE DIFFERENCE MODELS ON THREE COMPONENT SETS

In this section we investigate how the additive difference model can be axiomatized in the case the stimuli are composed of three component sets. Since the formal proofs of most lemmas and theorems are very similar to those of the two component case, we sketch them only briefly.
Let $A = \{a,b,c,d,\ldots\}$, $P = \{p,q,r,s,\ldots\}$

and $X = \{u, v, w, x, \ldots\}$ denote the three component sets. Then the set S of stimuli is given by $S = A \times P \times X = \{apu, bqv, cwx, \ldots\}$.
Let $\geq_o$ be the binary preference relation on S. We assume that $\geq_o$ satisfies the following axioms.

1. Axiom 1.

$\geq_o$ is a reflexive and connected relation. We let $>_o$ and $=_o$ denote the asymmetric and the symmetric part of $\geq_o$.

2. Axiom 2. Triple cancellation conditions.

(1) TCA-PX:

$$\left.\begin{array}{l} apu \geq_o bqv \\ brw \geq_o asx \\ csx \geq_o drw \end{array}\right\} \text{ imply } cpu \geq_o dqv$$

(2) TCP-AX:

$$\left.\begin{array}{l} apu \geq_o bqv \\ cqx \geq_o dpy \\ dry \geq_o csx \end{array}\right\} \text{ imply } aru \geq_o bsv$$

(3) TCX-AP:

$$\left.\begin{array}{l} apu \geq_o bqv \\ crv \geq_o dsu \\ dsx \geq_o cry \end{array}\right\} \text{ imply } apx \geq_o bqy$$

3. Axiom 3. Quintuple cancellation condition.

(1) QCA:

$$\text{if } \left\{\begin{array}{l} apx \geq_o bqv \\ b'qv \geq_o a'px \end{array}\right. \text{ and } \left\{\begin{array}{l} bp'x' \geq_o cq'v' \\ c'q'v' \geq_o b'p'x' \end{array}\right.$$

then $cry \geq_o asz$ implies $c'ry \geq_o a'sz$.

(2) QCP:

$$\text{if } \left\{\begin{array}{l} apx \geq_o bqy \\ bq'y \geq_o ap'x \end{array}\right. \text{ and } \left\{\begin{array}{l} a'qx' \geq_o b'ry' \\ b'r'y' \geq_o a'q'x' \end{array}\right.$$

then $crz \geq_o dpw$ implies $cr'z \geq_o dp'w$.

(3) QCX:

$$\text{if } \left\{\begin{array}{l} apu \geq_o bqv \\ bqv' \geq_o apu' \end{array}\right. \text{ and } \left\{\begin{array}{l} a'p'v \geq_o b'q'w \\ b'q'w' \geq_o a'p'v' \end{array}\right.$$

then $crw \geq_o dsu$ implies $crw' \geq_o dsu'$.

4. Solvability conditions.

For each component set we need unrestricted solvability for two unknowns and restricted solvability for one unknown. E.g. for the first component set A these solvability conditions can be formulated in the following way:

(1) USA: given $p,q \in P$ and $x,y \in X$, there exist $a,b \in A$ such that
$apx =_o bqy$.

(2) RSA: if $apx \geq_o bqy$ and $bqy \geq_o cpx$, there exist $d \in A$ such that
$dpx =_o bqy$.

Similar definitions can be stated for the component sets P and X.

5. Archimedean conditions.

The definition of a strictly bounded standard sequence for each component set can be adapted from the two component case described in the previous section. Subsequently, the archimedean conditions state that any strictly bounded standard sequence is finite.

We now investigate what kind of representation for $\geq_o$ is implied by these axioms.

In a way similar to what was done in the two component case, we define relations R_A on AxA, R_P on PxP and R_X on XxX:

(1) for arbitrary $a,b,c,d \in A$: $(a,b)R_A(c,d)$ iff there exist $p,q \in P$ and $x,y \in X$ such that $apx \geq_o bqy$ and $dqy \geq_o cpx$,

(2) for arbitrary $p,q,r,s \in P$: $(p,q)R_p(r,s)$ iff there exist $a,b \in A$ and $x,y \in X$ such that $apx \geq_o bqy$ and $bsy \geq_o arx$,

(3) for arbitrary $v,w,x,y \in X$: $(v,w)R_X(x,y)$ iff there exist $a,b \in A$ and $p,q \in P$ such that $apv \geq_o bqw$ and $bqy \geq_o apx$.

The relational structures $\langle AxA,R_A\rangle$, $\langle PxP,R_P\rangle$ and $\langle XxX,R_X\rangle$ all satisfy the axioms of an algebraic difference structure as formulated by Krantz et al. (1971). Since these proofs are very similar to the proofs of the corresponding theorems in the two component case, we omit them here.

It follows that there exist functions $f: A \rightarrow Re$, $g: P \rightarrow Re$ and $h: X \rightarrow Re$ (these functions being into the reals) such that:

(1) $(a,b)R_A(c,d)$ iff $f(a)-f(b) \geq f(c)-f(d)$

(2) $(p,q)R_P(r,s)$ iff $g(p)-g(q) \geq g(r)-g(s)$

(3) $(v,w)R_X(x,y)$ iff $h(v)-h(w) \geq h(x)-h(y)$.

We define:

$$\Omega_A = \{\alpha;\ \alpha = f(a)-f(b) \text{ for some } a,b \in A\}$$
$$\Omega_P = \{\pi;\ \pi = g(p)-g(q) \text{ for some } p,q \in P\}$$
$$\Omega_X = \{\xi;\ \xi = h(x)-h(y) \text{ for some } x,y \in X\}$$

Next we define a function $F: \Omega_A x \Omega_P \rightarrow \Omega_X$ as follows.

Suppose $apx =_o bqy$ with $\alpha = f(a)-f(b)$

$\pi = g(p)-g(q)$

and $\xi = h(y)-h(x)$.

Then $\xi = F(\alpha,\pi)$.

We can prove that F has the following properties (again, proofs are omitted):

1. F is well defined.
   Suppose $apx =_o bqy$ and $a'p'x' =_o b'q'y'$ with
   $f(a)-f(b) = f(a')-f(b')$ and $g(p)-g(q) = g(p')-g(q')$.
   Then we also have: $h(y)-h(x) = h(y')-h(x')$.
2. $F(-\alpha,-\pi) = -F(\alpha,\pi)$.
3. F is monotonically increasing in both arguments:
   (1) $\alpha \geq \alpha'$ iff $F(\alpha,\pi) \geq F(\alpha',\pi)$
   (2) $\pi \geq \pi'$ iff $F(\alpha,\pi) \geq F(\alpha,\pi')$.

Finally, we have the following

Representation theorem:

$$apx \geq_o bqy \text{ iff } F[f(a)-f(b),g(p)-g(q)] + h(x)-h(y) \geq 0.$$

This representation theorem shows that the axioms, which were formulated above, are not sufficient to ensure that the function F has an additive structure. Hence, at this moment we have not yet arrived at a complete axiomatization of the additive difference model on three component sets. In order to achieve this goal we need some supplementary conditions, which should imply that there exist functions ψ, ϕ_1 and ϕ_2 such that $F(\alpha,\pi) = \psi[\phi_1(\alpha)+\phi_2(\pi)]$.

The most obvious way to derive these supplementary conditions consists of defining a relation $\geq^*$ on $\Omega_A x \Omega_P$ as follows:

$$(\alpha,\pi) \geq^* (\alpha',\pi') \text{ iff } F(\alpha,\pi) \geq F(\alpha',\pi').$$

If the structure $\langle \Omega_A, \Omega_P, \geq^* \rangle$ satisfies the axioms of an additive conjoint measurement structure, the representation we have in view can be derived.

Now the fact that $\geq^*$ is a weak order and that it satisfies independence and unrestricted solvability merely follows from the properties of F. However, in order to ensure that $=^*$ satisfies the Thompsen condition we need an additional cancellation condition:

$$\text{if } \begin{cases} apu =_\circ bqv \\ a'p'u =_\circ b'q'v \end{cases} \text{ and } \begin{cases} a'p''w =_\circ b'q''x \\ a''pw =_\circ b''qx \end{cases}$$

then $ap''y =_\circ bq''z$ iff $a''py =_\circ b''qz$.

The proof that this additional cancellation axiom implies the Thompsen condition for $=*$ is straightforward.
From $apu =_\circ bqv$ and $a'p'u =_\circ b'q'v$ it follows that

$$F[f(a)-f(b),g(p)-g(q)] = F[f(a')-f(b'),g(p')-g(q')]$$

or say $F(\alpha,\pi) = F(\alpha',\pi')$.

From $a'p''w =_\circ b'q''x$ and $a''pw =_\circ b''qx$, it follows:

$$F[f(a')-f(b'),g(p'')-g(q'')] = F[f(a'')-f(b''),g(p)-g(q)]$$

or say $F(\alpha',\pi'') = F(\alpha'',\pi)$.

From $ap'y =_\circ bq''z$ and $a''py =_\circ b''qz$,

$$F[f(a)-f(b),g(p'')-g(q'')] = F[f(a'')-f(b''),g(p)-g(q)]$$

or say $F(\alpha,\pi'') = F(\alpha'',\pi)$.

Stated in terms of $=*$ this becomes:

if $(\alpha,\pi) =*(\alpha',\pi')$
and $(\alpha',\pi'') =*(\alpha'',\pi)$
then $(\alpha,\pi'') =*(\alpha'',\pi')$.

This is exactly the Thompsen condition for $=*$.

Finally, the structure $\langle \Omega_A,\Omega_P,\succsim*\rangle$ should also satisfy the archimedean condition for additive conjoint structures. By translating this condition in terms of the preference relation $\succsim_\circ$, we arrive at a stronger version of the archimedean axiom for the structure $\langle AxPxX, \succsim_\circ\rangle$ than originally formulated in Axiom 5. We will not dwell on this rather technical point here.

If $\langle AxPxX,\succsim_\circ\rangle$ satisfies the five axioms stated earlier together with the additional cancellation condition and the strong archimedean condition, then there exist functions ψ, ϕ_1 and ϕ_2 such that $apx \succsim_\circ bqy$ iff

$$\psi[\phi_1(f(a)-f(b)) + \phi_2(g(p)-g(q))] + h(x)-h(y) \geq 0$$

in which ψ is a monotonically increasing function of its argument. Then we also have: $apx \succsim_\circ bwy$ iff

$$\phi_1[f(a)-f(b)] + \phi_2[g(p)-g(q)] \geq \psi^{-1}[h(y)-h(x)] .$$

Putting $\phi_3(\xi) = -\psi^{-1}(-\xi)$, we have the final result: apx $\geq_o$ bqy iff

$$\phi_1[f(a)-f(b)] + \phi_2[g(p)-g(q)] + \phi_3[h(x)-h(y)] \geq 0$$

in which ϕ_3 is also an increasing function of its argument.

REFERENCES

[1] Beals, R., Krantz, D.H. & Tversky, A., The Foundations of Multidimensional Scaling, Psychological Review, 75 (1968), 127-142.

[2] Fishburn, P.C., Lexicographic Additive Differences, Journal of Mathematical Psychology 21 (1980), 191-218.

[3] Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A., Foundations of Measurement. Vol. I, New York (1971).

[4] Ng, Y.-K., Sub-semiorder: a Model of Multidimensional Choice with Preference Intransitivity, Journal of Mathematical Psychology 16 (1977), 51-59.

[5] Tversky, A., Intransitivity of Preferences, Psychological Review 76 (1969), 31-48.

[6] Van Acker, P., Models for Intransitive Choice, Doctoral Dissertation, University of Nijmegen (1977).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

GENERALIZATIONS OF INTERVAL ORDERS

Jean-Paul Doignon
Université Libre de Bruxelles

We survey some recent, yet unpublished results that generalize interval and semi orders in three directions: (i) multiple thresholds, (ii) partial structures, (iii) infinite underlying set.

1. INTRODUCTION

In order to modelize an individual's preferences, we consider a finite set X of objects on which is given an asymmetric relation P. This relation contains the ordered pairs ab such that b is strictly preferred to a. We also have on X the indifference relation I that we assume reflexive, symmetric and disjoint from P.

The classical "threshold model" of Luce (1956) will be generalized here. Recall the fundamental idea: one attributes a utility u(x) and a threshold t(x) to each object x from X in such a way that y is preferred to x when the utility of y sufficiently exceeds that of x, more precisely when $u(x) + t(x) < u(y)$. Two well known results stated below give necessary and sufficient conditions for the existence of such a representation of P; the threshold t(x) is either allowed to be dependent on x, or it has a constant value. In these results, the representation of the indifference comes from the assumption that among two objects the individual either prefers one to the other, or is indifferent between them. This assumption that we will later drop amounts to $I = X \times X - (P \cup P^{-1})$.

We will denote by R^{-1} the converse, by $\underline{R}$ the complementary and by R^0 the "converse of the complementary" relation of a given relation R, and by RS the relative product of the relations R and S.

THEOREM (Fishburn (1970)). There exist mappings $u: X \to Re$ and $t: X \to Re$ with $t(x) > 0$ such that

$$x\ P\ y \Leftrightarrow u(x) + t(x) < u(y),$$

$$x\ I\ y \Leftrightarrow \begin{cases} u(x) + t(x) \geq u(y) \\ u(y) + t(y) \geq u(x) \end{cases}$$

if and only if

$P \cup I$ is connected

and $PIP \subset P$ (or equivalently $PP^0P \subset P$).

THEOREM (Scott and Suppes (1958)). Moreover, t can be taken with a constant value if and only if

$IP^2 \subset P$ (or equivalently $P^0P^2 \subset P$).

Pairs (P,I) satisfying the first theorem are called interval orders; those satisfying the second theorem are called semi orders.

The conditions involving the utilities and thresholds can be formulated more geometrically. First, put $g = u$ and $f = u+t$. The irreflexivity of P implies $g \leq f$, and we associate to each object x from X the interval $[g(x), f(x)]$ on the real line. The preference P is interpreted as the precedence of such intervals, while the indifference corresponds to overlapping. Also the threshold becomes the interval length.

Noticing that

$$x\ P\ y \Leftrightarrow f(x) < g(y)$$

we come close to another well known result concerning this time the so-called Guttman scales (after e.g. Guttman (1944)).

THEOREM (Riguet (1948); Ducamp and Falmagne (1969)). Given a relation R on a finite set X, there exist two mappings f and g from X to Re satisfying

$$x\ R\ y \Leftrightarrow f(x) < g(y)$$

if and only if

$RR^0R \subset R$.

Relations satisfying the above conditions will be called biorders, after Doignon, Ducamp and Falmagne (1983+); Monjardet (1978), in his nice survey paper, calls them Ferrers relations.

Another useful characterization is in terms of the 0-1 matrices of the relation: R is a biorder iff for some simple ordering of the row elements and some simple ordering of the column elements, the resulting matrix has a step type form (see fig. 1).

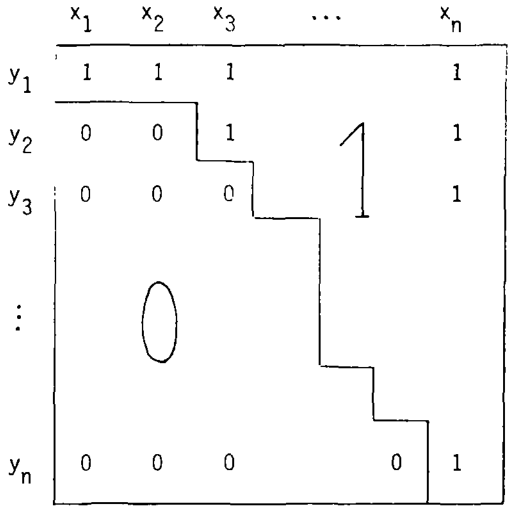


Figure 1

Fishburn's theorem on interval orders can be trivially deduced from the last theorem, because the irreflexive biorders are exactly the asymmetric parts P of interval orders. (Compare with the years and journals of publications!). For semi orders one can show that it is possible to use the same simple order on row and column elements when looking for a step type 0-1 matrix. We refer the reader to Monjardet (1978) and Roberts (1979) for this and many other related results, and now come to the promised survey of recent generalizations.

2. MULTIPLE THRESHOLDS

Suppose that we distinguish among the individual's preferences several levels. We then form for each of the levels a relation P_i on the set X of objects, and want to obtain general conditions for the existence of a representation for the relations I, P_1, P_2, ..., P_k using a utility function together with one threshold function for each level of preference. (Thus in particular each P_i will be the asymmetric part of some interval order).

As in the classical case, the notion of biorder appears as a useful tool. Without entering into details, let us say that a careful study of the construction of the simple orders leading to a step type matrix easily gives the generalization to multiple simultaneous biorders. Here is an example of the results one can obtain.

PROPOSITION

Given relations $R_1, R_2, \ldots, R_k$ on the finite set X, the three following conditions are equivalent:

(i) some simple order on X for the row elements together with some simple order on X for the column elements give simultaneous step type matrices for $R_1, R_2, \ldots, R_k$;

(ii) there exist mappings $f_1, f_2, \ldots, f_k, g$ from X to Re satisfying:

$$x \, R_i \, y \Leftrightarrow f_i(x) < g(y)$$

and $f_i(x) < f_i(y) \Rightarrow f_j(x) \leq f_j(y)$;

(iii) the relations

$$\underline{R}_1 R_1^{-1} \cup \underline{R}_2 R_2^{-1} \cup \ldots \cup \underline{R}_k R_k^{-1}$$

and $R_1^{-1} \underline{R}_1 \cup R_2^{-1} \underline{R}_2 \cup \ldots \cup R_k^{-1} \underline{R}_k$

are asymmetric.

The relations involved in this proposition thus admit a pictorial presentation examplified in Figure 2.

It is interesting to notice that one can use the same simple order for row and column elements if and only if the union of the two relations appearing in (iii) above is itself asymmetric.

As a corollary, we obtain a representation theorem for multithreshold preference relations.

COROLLARY.

Given a symmetric relation I and asymmetric relations $P_1, P_2, \ldots, P_k$ on the finite set X, there exist mappings $u, t_1, t_2, \ldots, t_k$ from X to Re with $t_i(x) > 0$ such that

$$x \, I \, y \Leftrightarrow \begin{cases} u(x) + t_k(x) \geq u(y) \\ u(y) + t_k(y) \geq u(x), \end{cases}$$

$$x\, P_i\, y \Leftrightarrow u(x) + t_i(x) < u(y),$$
$$t_1 \geq t_2 \geq \ldots \geq t_k,$$

and

$$u(x)+t_i(x) < u(y)+t_i(y) \Rightarrow u(x)+t_j(x) \leq u(y)+t_j(y)$$

if and only if

$$P_1 \subset P_2 \subset \ldots \subset P_k,$$

$$I \cap P_k = \emptyset,$$
$$I \cup P_k \text{ is connected,}$$

and the relations

$$\underline{P}_1 P_1^{-1} \cup \underline{P}_2 P_2^{-1} \cup \ldots \cup \underline{P}_k P_k^{-1}$$

and

$$P_1^{-1} \underline{P}_1 \cup P_2^{-1} \underline{P}_2 \cup \ldots \cup P_k^{-1} \underline{P}_k$$

are symmetric.

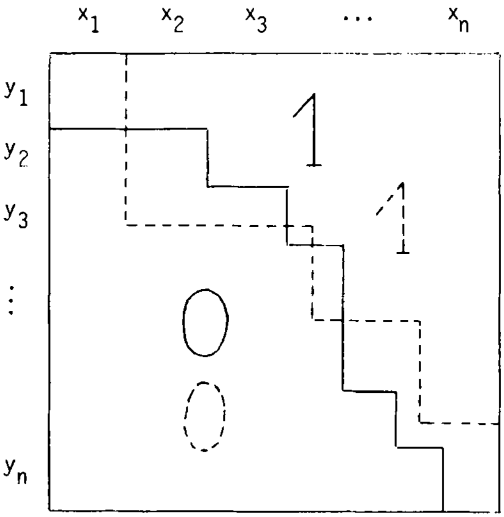


Figure 2

Many other results can be established as easily as this one (for instance, similar statements obtained by dropping some of the conditions). Nevertheless there remain open problems, and we present one very general question.

PROBLEM.

Find conditions on I, P_1, P_2, ..., P_k that ensure the existence of a representation using u, t_1, t_2, ..., t_k with for some prescribed indices i, the threshold t_i being constant (i.e. independent of the object x).

2. PARTIAL STRUCTURES

We consider again a pair (P,I) of preference and indifference relations on X, with P asymmetric and I symmetric and reflexive, but we do not assume that any pair of two objects is in one of these two relations. There are many motivations for studying this situation; for instance, the individual may consider that two objects are incomparable, or else he considers that they are comparable but we do not have the complete information on his judgements. These motivations lead Roy (1980) to consider numeric representations with the double implication replaced by a single one. More precisely, one looks again for mappings u and t, satisfying this time:

$$x \, P \, y \Rightarrow u(x) + t(x) < u(y)$$

$$\text{and} \quad x \, I \, y \Rightarrow \begin{cases} u(x) + t(x) \geq u(y) \\ u(y) + t(y) \geq u(x). \end{cases}$$

Historically, the first result is due to Roy (1980) and involves a constant threshold. Then Roubens and Vincke (1983+) settled the variable threshold case.

THEOREM (Roy (1980)).

There exist mappings u and t as above, with the threshold t(x) being constant, if and only if each cycle of the relation $P \cup I$ has more pairs from I than from P.

THEOREM (Roubens and Vincke (1983+)).

There exist mappings u and t as above if and only if the relation PI is acyclic.

We have obtained generalizations of these two theorems involving the notion of biorder. More precisely, a biorder is used to separate two relations R and S; the numerical formulation is then easily derived. Notice that the theorems are particular cases with P = R and I = S^{-1}.

PROPOSITION.

Given relations R and S on the finite set X, the following conditions are equivalent:

(i) there exist mappings f and g from X to Re satisfying

$$x\,R\,y \Rightarrow f(x) < g(y),$$
$$x\,S\,y \Rightarrow f(x) \geq g(y),$$

and $g(x) \leq g(y)$ and $f(y) \leq f(x) \Rightarrow x = y$;

(ii) there exists a biorder B such that

$R \subset B$ and $S \subset \underline{B}$,

and $B^0B^2 \subset B$;

(iii) in each cycle of $R \cup S^{-1}$ there are strictly more pairs from S^{-1} than from R.

Remark that the third part of condition (i) in fact avoids inclusion between some intervals; using the finiteness of X and assuming R irreflexive it can be shown that (i) is equivalent to the possibility of choosing f and g in such a way that those intervals have constant length.

PROPOSITION.

Given relations R and S on the finite set X, the following conditions are equivalent:

(i) there exist mappings f and g from X to Re satisfying

$$x\,R\,y \Rightarrow f(x) < g(y),$$

and $x\,S\,y \Rightarrow f(x) \geq g(y)$;

(ii) there exist a biorder B such that

$R \subset B$ and $S \subset \underline{B}$;

(iii) RS^{-1} is acyclic.

These propositions also lead to interesting sharpenings of the theorems due to Roy and Roubens and Vincke (a manuscript with A. Ducamp and J.-C. Falmagne is in preparation). For instance, one asks for a description of the pairs xy such that, for any mappings u and t, holds the relation $u(x) + t(x) < u(y)$ (these are the pairs that must be added to the observed preferences in order to get a threshold model). Here is the answer: those pairs form the relation $P \cup (PI)^*P$, where * symbolizes the transitive closure operator.

3. INFINITE CASE

We have assumed for convenience that the set X of objects is finite. But in fact most of our results also hold for infinite sets X, under the condition that the ordered set Re be replaced by some simply ordered set E (i.e. one only states the existence of E and of the mappings from X to E).

More difficult is the problem involving an infinite set X and mappings from X to Re; density conditions will be needed. For instance, Fishburn (1973) gives somewhat intricated conditions in the case of an interval order. Simpler conditions were recently forged, that apply as well for biorders. (Recall that the interval orders are the irreflexive biorders).

PROPOSITION (Doignon, Ducamp and Falmagne (1983+)).

For a given relation R on a set X, there exist mappings f and g from X to Re such that

$$x \; R \; y \;\Leftrightarrow\; f(x) < g(y)$$

if and only if there exist a countable subset M of X which is strictly dense, that is:

for any x and y in X, there exists m in M such that:

$x \; R \; m$ and not $m \; R^{-1}\underline{R} \; y$,

or $m \; R \; y$ and not $x \; \underline{R}R^{-1} \; m$.

4. CONCLUSIONS

The results in this brief survey describe conditions for the existence of numerical representations of a certain type. Our generalizations (often taken from joint work with A. Ducamp and J.-C. Falmagne) are based on the notion of biorder. From our point of view, this notion clarifies the subject because it synthetizes different aspects of it, and provides the way to simple and general proofs. As a conclusion, let us propose a new maxim: if you are looking for a threshold representation theorem, first find where the biorder is!

REFERENCES

[1] Doignon, J.-P., Ducamp, A., and Falmagne, J.-C., On realizable biorders and the biorder dimension of a relation, Journal of Math. Psychol., to appear.

[2] Ducamp, A., and Falmagne, J.-C., Composite measurement, Journal of Math. Psychol., 6 (1969), 359-390.

[3] Fishburn, P.C., Intransitive indifference with unequal indifference intervals, Journal of Math. Psychol., 7 (1970), 144-149.

[4] Fishburn, P.C., Interval representations for interval orders and semiorders, Journal of Math. Psychol., 10 (1973), 91-105.

[5] Guttman, L., A basis for scaling quatitative data, Amer. Soc. Rev., 9 (1944), 139-150.

[6] Luce, R.D., Semi-orders and a theory of utility discrimination, Econometrica, 24 (1956), 178-191.

[7] Monjardet, B., Axiomatiques et propriétés des quasi-ordres, Math. Sci. Humaines, 63 (1978), 51-82.

[8] Riguet, J., Les relations de Ferrers, C.R. Acad. Sci. Paris, 232 (1951), 1729-1730.

[9] Roberts, F.S., Measurement theory, Encyclopedia of Mathematics and its Applications, Vol. 7, Addison-Wesley, Reading, Mass. (1979).

[10] Roubens, M., and Vincke, Ph., On partial interval orders, to appear.

[11] Roy, B., Préférence, indifférence, incomparabilité, Documents du LAMSADE, 9, Université de Paris-Dauphine (1980).

[12] Scott, D., and Suppes, P., Foundational aspects of theories of measurement, Journal Symbolic Logic, 23 (1958), 113-128.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B. V. (North-Holland), 1984

RESEAUX SOCIAUX GENERALISES:
EN COMBINANT GRAPHES ET HYPERGRAPHES

Claude Flament

Département de Psychologie
Université de Provence

Centre de Mathématique Sociale
E.H.E.S.S., Marseille

Nous évoquons ici quelques notions en relation avec les réseaux sociaux, qui combinent graphes et hypergraphes sur un même ensemble d'individus, et en ajoutant un ensemble de traits caractérisant ces individus. Nous traiterons principalement le système de caractérisation.

1. INTRODUCTION

Traditionnellement, un réseau social est représenté par un graphe, ou plusieurs, dont les sommets correspondent aux individus. Plus récemment, divers travaux proposent de considérer qu'un réseau social est en fait un système représentable par un hypergraphe (Berge (1970)), c'est-à-dire, une famille de sous-ensembles de l'ensemble des individus : groupes d'affinité, associations, membres d'une même profession, classes d'individus ayant des caractérisations équivalentes... (Degenne (1981), reprenant une idée de Bouglé (1897), Seidman (1981), Doreian (1981), Bonacich et Domhoff (1981)).

L'intérêt de cette nouvelle approche ne doit pas faire oublier les apports de l'ancienne. Aussi nous évoquons ici quelques notions qui combinent graphes et hypergraphes sur un même ensemble d'individus. On y ajoute un ensemble de traits caractérisant ces individus. Un tel système peut être considéré comme un réseau social généralisé.

Pour la simplicité, nous considèrerons d'abord séparément : graphe et hypergraphe sur l'ensemble X des individus, d'une part; et d'autre part, l'hypergraphe sur X, un hypergraphe sur un ensemble Y de traits et une relation de caractérisation entre X et Y. Mais, mathématiquement, le premier sous-système est un cas particulier du second : on retrouve le premier en faisant $X = Y$, et en identifiant les deux hypergraphes. Nous traiterons donc principalement le système de caractérisation.

2. NOTATIONS

Un hypergraphe peut être considéré comme un graphe biparti : de l'ensemble des sommets vers l'ensemble des arêtes.

Nous désignerons par la même lettre un graphe, la relation qu'il représente, et la matrice qui lui est associée. Ainsi : $(xy) \in G$, xGy, $(xy)=1$ dans G, sont des formulations équivalentes.

La composition des graphes, ou des relations, s'écrit comme le produit booléen des matrices. On a : $xGHy$ si et seulement si existe un z tel que : xGz et zHy.

Nous désignons par G' la matrice transposée de la matrice G, ainsi que graphe et relation correspondants.

Soit deux ensembles X et Y; sur X, un hypergraphe H; sur Y, un hypergraphe K; entre X et Y, un graphe biparti R. Un tel système sera noté $[H,R,K]$.

Si $x_i HH' x_j$, c'est que les individus x_i et x_j sont dans au moins un même sous-groupe du réseau. De même, si $y_i KK' y_j$, c'est que les traits y_i et y_j sont dans au moins une même classe de caractérisation.

Un graphe de similitude est un graphe réflexif et symétrique : HH' et KK' sont de tels graphes. Si un graphe S de similitude est de plus transitif (i.e. : $SS=S$), c'est alors le graphe d'une équivalence.

3. PLENITUDE D'UN SYSTEME

Définition : Un système $[H,R,K]$ est :

- plein à gauche si : $R = HH'R$;
- plein à droite si : $R = RKK'$;
- plein s'il est plein à gauche et à droite.

La terminologie est justifiée par les résultats suivants, qui généralisent ceux de Riblier (1978).

Théorème : Si un système [H,R,K] est plein à gauche, (xHH'x') entraîne que x et x' ont exactement les mêmes successeurs dans R; s'il est plein à droite, (yKK'y') entraîne que y et y' ont exactement les mêmes prédécesseurs dans R.

Démonstration : Si R=HH'R :

(xHH'x' et x'Ry) ⇒ (xHH'Ry) ⇒ (xRy).

Commentaires : On voit donc que si une flèche R part d'une arête de H, il en part, dans la même direction, de chacun des sommets de cette arête. Symétriquement, si le système est plein à droite et si une flèche R arrive dans une arête de K, tous les sommets de cette arête reçoivent une flèche de la même origine.

Si le système est plein, et qu'on extrait de la matrice R la sous-matrice croisant une arête de H est une arête de K, on voit que ce bloc est totalement composé soit de 1, soit de zéros.

En remarquant que deux arêtes de H (ou de K) peuvent avoir une intersection non vide, on voit que les propriétés "mêmes successeurs" et "mêmes prédécesseurs" s'étendent aux classes des équivalences correspondant aux composantes connexes de H et de K.

Si nous sommes dans le cas d'un système plein, où X=Y et H=K, on voit que l'équivalence correspondant aux composantes connexes de H n'est autre que l'équivalence structurale de Lorrain et White (1971).

Or, bien des auteurs (par exemple, Sailer (1978)) pensent que cette notion est trop forte pour être utile dans l'étude des réseaux. C'est pourquoi nous proposons une notion plus faible mais plus utile. Mais avant, donnons, sans démonstration, un résultat facile à établir .

Proposition : Un système [H,R,K] est plein si et seulement si : R=HH'RKK'.

4. REGULARITE D'UN SYSTEME

Définition : Un système [H,R,K] est :

- régulier à gauche si : $RKK' = HH'RKK'$;
- régulier à droite si : $HH'R = HH'RKK'$;
- régulier s'il est régulier à droite et à gauche.

Théorème : Soit un système [H,R,K] ; considérons la sous-matrice de R croisant une arête de H et une arête de K; si le système est régulier à gauche, cette sous-matrice comporte, soit seulement des zéros, soit au moins un 1 dans chaque ligne; si le système est régulier à droite, cette sous-matrice comporte, soit seulement des zéros, soit au moins un 1 dans chaque colonne.

Démonstration : Si $RKK' = HH'RKK'$:
$(xHH'x' \text{ et } x'RKK'y) \Rightarrow (xHH'RKK'y) \Rightarrow (xRKK'y) \Rightarrow (\exists y':xRy')$.

Proposition : Si HH' ou KK' est une équivalence, le système [H,R,K] est régulier si et seulement si : $RKK' = HH'R$.

Démonstration : Il suffit de démontrer que $RKK' = HH'R$ entraîne que RKK' ou HH'R égale HH'RKK'; supposons que KK' est une équivalence, dont la transitivité s'écrit : $KK'KK' = KK'$; alors : $HH'RKK' = RKK'KK' = RKK'$.

Commentaires : Supposons que nous soyons dans le cas d'un système de caractérisation : les traits constituant une arête de K sont équivalents (relativement au problème traité), et un individu, pour avoir la caractérisation globale correspondant à une arête de K, n'a pas besoin d'en avoir tous les traits : il suffit qu'il ait tel trait OU tel autre trait de la classe. On voit donc que la régularité traduit un modèle disjonctif de classification.

Supposons maintenant que X = Y et H = K; par exemple, X est un ensemble de scientifiques d'un même secteur, et R est le graphe des citations. Si une arête représente une école particulière, on pourra juger qu'un individu s'intéresse à cette école s'il cite au moins l'un de ses membres; il n'est pas nécessaire qu'il les cite tous.

On pourrait développer dans le sens de la théorie structurale des rôles, chère à Harrison White et son équipe (White et al., (1976), Boorman et White (1976)).

Enfin (et ces problèmes sont tous liés, chez White comme chez nous), nous considérons que la notion de bloc régulier (sous-matrice de R avec au moins un 1 dans chaque ligne et chaque colonne) peut être mise à la base d'un nouveau "modèle de bloc" ("Blockmodel"; cf. : Arabie, Boorman et Levitt (1978), Arabie et Boorman (1982)). Nous avions exposé cette idée (dans le cas où HH' et KK' sont des équivalences) à la 9ème réunion du Groupe Européen de Psychologie Mathématique (Nijmegen (1978)).

5. PERSPECTIVES

Nous pouvons revenir aux réseaux sociaux généralisés : un ensemble X d'individus, et sur cet ensemble, une famille de graphes R_i, une famille d'hypergraphes H_j; un ou plusieurs ensembles de caractérisation Y_k, avec des hypergraphes K_k, et des relations Q_k de X vers Y_k.

Cette multiplicité d'ensembles, de graphes et d'hypergraphes offre d'intéressantes possibilités, dans la mesure où la régularité se conserve à travers plusieurs des opérations possibles sur ces graphes et hypergraphes, comme le montrent les propriétés suivantes, faciles à démontrer si les similitudes associées aux hypergraphes sont des équivalences :

- soit une famille de systèmes réguliers $[H, R_i, K]$ sur X et Y; le système $[H, \cup R_i, K]$ est régulier;
- soit une famille de systèmes réguliers $[H_i, R, K_i]$ sur X et Y; le système $[\cup H_i, R, \cup K_i]$ est régulier.

(Note : si on remplace l'union par l'intersection, les deux propriétés ne sont plus vraies.)

- soit deux systèmes réguliers, $[H, R, K]$ sur X et Y, et $[K, Q, L]$ sur Y et Z (X, Y et Z étant des ensembles distincts ou non); le système, sur X et Z, $[H, RQ, L]$ est régulier.

Cette propriété assure le maintien de la régularité à travers les compositions que propose Lorrain (1975). Mais on peut aussi considérer certains rapports entre systèmes, qui engendrent de la régularité; par exemple :

- soit, sur X et Y, deux systèmes [H, R_i, K] avec i = 1 ou 2, tels que: $HH'R_1 = R_2KK'$; alors, le système [H, R_1, K] est régulier à droite, et le système [H, R_2, K] est régulier à gauche.

6. PROBLEMES

Supposons une relation R entre X et Y; on veut trouver un hypergraphe H sur X, un hypergraphe K sur Y, tels que le système [H, R, K] soit régulier. Ce problème n'est pas résolu. Tout montre que la résolution du problème dans le cas simple évoqué, s'étendrait sans difficulté de principe au cas d'un réseau généralisé (nombreux ensembles, nombreux graphes).

Une solution combinatoire existe si on accepte que les hypergraphes recherchés correspondent à des équivalences. Mais l'application à des données réelles montre que presque toujours, les solutions exactes sont triviales, alors qu'on trouve (par tâtonnements) des solutions approchées très intéressantes. Il faudrait donc trouver une méthode, mêlant statistique et combinatoire, permettant de cerner la quasi-régularité (bloc avec au moins un 1 dans presque chaque ligne et dans presque chaque colonne).

Remerciements: Bien des idées de ce travail sont dues à Alain Degenne; toutes se sont développées en interaction constante avec son propre travail.

REFERENCES

[1] Arabie, P., Boorman, S.A., Levitt, P.R., Constructing blockmodels: how and why, Journal of Mathematical Psychology, 17 (1978), 21-63.

[2] Arabie, P., Boorman, S.A., Blockmodels: developments and prospects, in H.C. Hudson (Ed.), Classifying social data, San Francisco, Jossey-Bass, (1982), 177-259.

[3] Berge, C., Graphes et Hypergraphes, Paris, Dunod (1970).

[4] Bonacich, P., Domhoff, W., Latent classes and group membership, Social Networks, 3 (1981), 175-196.

[5] Boorman, S.A., White, H.C., Social structures from multiple networks, II, Role structures, American Journal of Sociology, 81 (1976), 1384-1446.

[6] Bougle, C., Qu'est-ce que la sociologie, Revue de Paris (1897).

[7] Degenne, A., Les réseaux de coopération et d'échange, Centre d'Analyse et de Mathématique Sociale, E.H.E.S.S., Marseille.

[8] Doreian, P., Polyhedral dynamics and conflict mobilization in social networks, Social Networks, 3 (1981), 107-116.

[9] Lorrain, F., White, H.C., Structural equivalence of individuals in social networks, Journal of Mathematical Sociology, 1 (1971), 49-80.

[10] Lorrain, F., Réseaux sociaux et classifications sociales, Paris, Hermann (1975).

[11] Riblier, W., Quelques développements de la théorie des graphes et applications à la prospective par analyse de système, Thèse de mathématiques economiques, Université Pierre et Marie Curie, Paris (1978).

[12] Sailer, L.D., Structural equivalence: meaning and definition, Computation and Application, Social Networks, 1 (1978), 73-90.

[13] Seidman, S.B., Structures induced by collections of subsets: a hypergraph approach, Mathematical Social Sciences, 1 (1981), 381-396.

[14] White, H.C., Boorman, S.A., Breiger, R.L., Social structure from multiple networks, I, Blockmodel of roles and positions, American Journal of Sociology, 81 (1976), 730-780.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

ON CONCEPTS OF THE DIMENSION OF A RELATION AND A GENERALIZED RECOGNITION EXPERIMENT

Klaus Herbst
Universität Regensburg

Confusion among stimuli is generalized to a binary relation R from the set of stimuli to its power set. If R induces a semimodular lattice of finite length, the concept of lattice dimension is meaningful. In any case R is the intersection of at least one set of biorders. The minimum number of biorders needed defines the biorder dimension of R. Data from an experiment with Morse codes are analysed.

1. INTRODUCTION

Consider the following recognition experiment. Two elements from a set A of stimuli are presented in sequence. A subject decides whether the first stimulus has been repeated or not. The experimental design is A x A. Let the data induce a binary relation S on A. This empirical relation S is often found to be non-symmetric like the confusion matrices which arise from repetitions of the experiment.

Multidimensional scaling procedures use confusion matrices in order to map the set of stimuli into a metric space such that the dissimilarity between pairs of stimuli is represented by a distance function. The foundation of these scalings is unsatisfactory with respect to the choice of the metric space and its dimension.

Beals, Krantz, and Tversky (1968) first gave a solution of the representation problem for a specific class of metric spaces. But they presuppose that the dimension of the structure is known in advance.

In this paper two concepts of the dimension of a relation are discussed, the lattice dimension and the biorder dimension.

2. A GENERALIZED RECOGNITION EXPERIMENT

The analysis relies on the following generalized recognition experiment: like in the preceding design a single stimulus $a \in A$ is presented at first. But now it is followed by a random sequence (b,d,c) of elements from A. The subject decides whether the first stimulus a has been repeated in the sequence (b,d,c). The total power set $P(A)$ is used such that the design of this experiment is $A \times P(A)$. Let the data induce a relation $R \subseteq A \times P(A)$.

3. LATTICE DIMENSION

For short let now P denote the power set of A. For any $p \in P$ we set $Rp = \{a \in A \mid aRp\}$ and define a relation on P, $R_P \subseteq P \times P$, by

$$pR_Pq \quad \text{iff} \quad Rp \subseteq Rq.$$

This allows us to introduce an *equivalence relation* on P, such that $p \sim q$ iff $Rp = Rq$, and to reduce P to a set of representatives of equivalence classes. R_P is called *irreducible* iff there exist no two equivalent elements p and q.

Any irreducible relation R_P fulfills the definition of a *partial order*. Moreover, it is a *lattice* provided that every pair $p,q \in P$ has a supremum and an infimum in R_P.

DEFINITION

R_P is a *lattice* iff

A1: R_P is a partial order

A2: For every $p,q \in P$ exists a *least upper bound* $r \in P$, $r = \sup(p,q)$. (supremum)

A3: For every $p,q \in P$ exists a *greatest lower bound* $s \in P$, $s = \inf(p,q)$. (infimum)

The concept of lattice dimension is based on the notion of a chain from the least to the greatest element in R_P which has maximal length. If all maximal chains in a lattice R_P have the same length m, we call m the *lattice dimension* of R_P, $\text{Ldim}\, R_P = m$.

A sufficient condition for the existence of $\text{Ldim}\, R_P$ can be formulated using the following definition.

DEFINITION

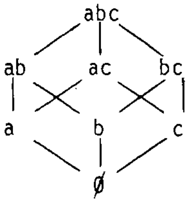
Let p,q,r be elements of P and $\bar{R}_p$ denote the negation of R_p. Then p covers q iff qR_pp and for all r holds $q\bar{R}_pr \vee r\bar{R}_pp$.

For any lattice R_p exists a dimension Ldim R_p if at least one of the following axioms holds:

A4: If p and q cover $\inf(p,q)$, then $\sup(p,q)$ covers p and q.
(upper semimodularity)

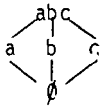
A5: If $\sup(p,q)$ covers p and q, then p and q cover $\inf(p,q)$.
(lower semimodularity).

If a subject in a generalized recognition experiment commits no errors, which means that no confusion happens, then R_p is isomorphic to a Boolean lattice of sets $\langle \mathcal{P}(A), \subseteq \rangle$ consisting of the power set of stimuli and the relation of set inclusion. The lattice dimension is in this case equal to the cardinality of A, Ldim $R_p = |A|$. The following Hasse diagram represents an example with $|A| = 3$.



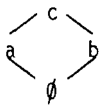
However, if confusion occurs, R_p may still be a lattice of some dimension smaller than $|A|$. Here are two examples with $|A| = 3$. By ab... we denote just that $p \in P = \mathcal{P}(A)$ which consists of the elements a,b,... of A.

Example 1: aRbc, bRac and cRab
induces ab $\sim$ ac $\sim$ bc $\sim$ abc.
R_p reduces to



and Ldim $R_p = 2$.

Example 2: If aRc, bRc and cRab,
R_p reduces to



and Ldim $R_p = 2$.

Notice, that in example 2 the confusion is non-symmetric so that the stimulus c dominates the stimuli a and b.

This asymmetry is preserved in a representation of $\langle P,R_p\rangle$ in a *vectorspace*. If A1-A5 are valid and two supplementary axioms, A6 and A7 listed below, are satisfied then $\langle P,R_p\rangle$ is homomorphic to a vectorspace V over a field F (Birkhoff (1967), Drösler (1980)). The elements of P correspond to subspaces of V so that the single stimulus c in example two would be represented not by a vector but by a plane in V.

A6: Every chain in R_p consists of a *finite* number of elements.

A7: Let $\emptyset$ be the least and e the greatest element in R_p. For each $p \in P$ exists a *complement* $q \in P$ such that $\inf(p,q) = \emptyset$ and $\sup(p,q) = e$.

4. EMPIRICAL EXAMPLE

A generalized recognition experiment was run with five subjects and the following six Morse Codes (Herbst (1981)).

a	. . - .
b	. - - .
c	. - . -
d	- . . -
e	- . - -
f	- - . .

With $|A| = 6$ and $|P(A)| = 64$ the design $A \times P(A)$ has the order 6×64.

Typically the subject heard at first a single code and afterwards a sequence of codes both chosen at random. Then it decided whether the single code had been repeated in the sequence.

The relation $R \subseteq A \times P$ was operationally defined by aRp iff the frequency of yes-answers exceeds the frequency of no-answers. The results of this experiment may be summarized as follows.

In no case the concept of lattice dimension can be applied. Either axiom A2 (supremum) is not fulfilled or not all maximal chains have the same length. The axioms A4 and A5 (semimodularity) do not hold. The concept of a lattice structure is too restrictive for these data.

We observe that subjects seldom miss a signal, that means if $a \in p$, aRp holds. But false alarms are numerous, we find aRp although $a \notin p$. Therefore R_p can be regarded as an extension of a Boolean relation of sets. Consequently, R is reducible insofar as equivalence classes can be formed in P but not in A. The Cartesian product $A \times P$ can be reduced from 6×64

to 6 × 35 at most and 6 × 12 at least. Subsets of A which are not ordered in a Boolean relation of sets are ordered in the empirical relations generated by the five subjects. This leads to the following idea.

5. BIORDER DIMENSION

In an extreme case of order R would reduce to a triangle

$$\begin{array}{c|cccc|} & & P & & \\ \hline & R & & & \\ A & R & R & & \\ & R & R & R & \\ & R & R & R & R \\ \hline \end{array}$$

that means R_P and a correspondingly defined R_A are connected, i.e. they are weak orders.

In this case R fulfills the definition of a biorder.

DEFINITION

Let $R \subseteq A \times P$. R is a biorder R iff for all $a,b \in A$ and $p,q \in P$, aRp and pRq implies aRq or bRp.

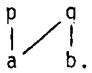
A representation of A and P in $\mathbb{R}$, the set of real numbers, assures the following theorem.

THEOREM

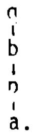
If R is a biorder, there exist two mappings $f\colon A \to \mathbb{R}$ and $g\colon P \to \mathbb{R}$, such that for all $a \in A$ and $p \in P$ holds aRp iff $f(a) \leq g(p)$.

The basic subrelation of a biorder R is

$$\begin{array}{c|cc|} & p & q \\ \hline a & R & R \\ b & & R \\ \hline \end{array}$$

with the diagram 

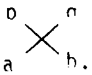
Doignon, Ducamp and Falmagne (to appear) show that if R is a biorder, a relation $Q \subseteq (A \cup P) \times (A \cup P)$ can be defined which has the property of a weak order. Its representation in $\mathbb{R}$ is the well-known Guttman scale.

The corresponding diagram of the basic example just shown is 

The basic forbidden subrelation is

$$\begin{array}{c|cc|} & p & q \\ \hline a & R & \\ b & & R \\ \hline \end{array}$$

the complement of which, $\bar{R}$ is illustrated by:



As could be expected, no subject realized a biorder. Unidimensional representations are impossible. The matrix below shows a typical relation R. Notice the number of equivalence classes and the frequency of the forbidden subrelation.

Example: relation R of subject nº 4.

A

		a	b	c	d	e	f	
	∅							
	a	R						
	b	R	R					∿ ab
	c			R				
	d				R	R		∿ e, de
	f						R	
	ac	R		R				
	ad	R			R	R		∿ ae, ade
	af	R					R	∿ abf
	bc	R	R	R				∿ abc
	bd	R	R		R			
	be	R	R		R	R		∿ abd, abe, bde, abde
	bf	R	R				R	
	cd			R	R			
	ce			R	R	R		∿ cde
	cf			R			R	
	df				R	R	R	∿ ef, def
P	acd	R		R	R			
	ace	R	R	R	R	R		∿ bcd, bce, abcd, abce, bcde
	acf	R		R			R	
	adf	R			R	R	R	∿ adef
	aef	R	R		R	R	R	∿ bef, abef, bdef, abdef
	bcf	R	R	R			R	∿ abcf
	bdf		R		R	R	R	
	cdf			R	R	R	R	∿ cdef
	cef			R		R	R	
	abdf	R	R		R		R	
	acde	R		R	R	R		
	acdf	R	R	R	R		R	∿ bcdf
	acef	R		R	R	R	R	
	bcef	R	R	R	R	R	R	∿ abcde, abcdf, abcef, acdef, bcdef, abcdef

The concept of biorder dimension or bidimension of a relation relies on the following theorem.

THEOREM

For any relation $R \subseteq A \times P$ exist two families of biorders $\{B_i\}$ and $\{B'_j\}$ such that $R = \cap B_i = \cup B'_j$.

We restrict the notion of bidimension to the intersection case.

DEFINITION

The bidimension of a relation $R \subseteq A \times P$, Bidim R, is the smallest cardinal n, such that there exists a collection of biorders $\{B_i \mid i \in I\}$ with $|I| = n$ and $R = \underset{i}{\cap} B_i$.

The multidimensional representation theorem has been proven by Doignon, Ducamp and Falmagne (to appear).

THEOREM

Let R be a countable relation and Bidim R = n, then there exist two mappings $f: A \to \mathbb{R}^n$ and $g: P \to \mathbb{R}^n$, such that aRp iff $f_i(a) \leq g_i(p)$, for all components $i=1,\dots,n$ of the vectors $f(a), g(p) \in \mathbb{R}^n$.

In a generalized recognition experiment we might expect the subject to give a positive answer exactly if all stored and highly evaluated features of the foregoing single stimulus are judged to be included in the following sequence of stimuli.

Bidim R may be interpreted as the conjoint dimension of the confusability of the single stimuli on the one hand and of the power of the subsets of stimuli to lead to confusion on the other hand.

The evaluation of the bidimension of a relation is a problem which often turns out to be hard.
If R is irreducible, that means R_P an R_A are irreducible, an upper bound is given by Bidim $R \leq \min(|A|,|P|)$. For $P = P(A)$ we have Bidim $R \leq |A|$, the cardinality of the set of stimuli. In order to give a lower bound, the following definition is useful.

DEFINITION

Let $A' \subseteq A$ and $P' \subseteq P$. Then $K \subseteq A' \times P'$ is called a crown iff the complement of K, denoted by $\bar{K}$, is a one to one correspondence between A' and P'.

In a crown K exists for each pair $a,b \in A'$ exactly one pair $p,q \in P'$ such that $(a,b) \times (p,q)$ is a forbidden subrelation. This leads to a standard result.

LEMMA

The bidimension of a crown $K \subseteq A' \times P'$ is equal to the cardinality of its domain and range, $\text{Bidim } K = |A'| = |P'|$.

Now, let $K_{max} \subseteq R$ be the greatest crown in R such that for all crowns $K_j \subseteq R$ holds $|K_{max}| \geq |K_j|$, and let R be irreducible. Then, in the special case considered here, we have the inequality

$$\text{Bidim } K_{max} \leq \text{Bidim } R \leq |A|.$$

If Bidim $R > 2$ the evaluation of the bidimension is called a NP-complete problem. That means, the computing time to check whether Bidim $R > n > 2$ is an exponential function of the inherent conditions given by R (Yannakakis (to appear)). Doignon, Ducamp and Falmagne (to appear) formulate these conditions in the language of graph theory. In principle, the elements of $\bar{R}$ (the complement of R) are regarded as vertices and special sequences in $\bar{R}$ as edges of a hypergraph H(R). These sequences arise from generalizations of the basic forbidden subrelation. Now, the vertices of H(R) are coloured by a family of colours $\{C_i \mid i \in I\}$ such that no edge is monochromatic, i.e. is coloured by one colour only.

The minimum number of colours I which allow such a coloration, is called the chromatic number Chrom H(R). It can be shown that Bidim R = Chrom H(R). The main idea of the proof may be sketched as follows.
If Bidim R = Chrom H(R) = n, each coloration $C_i \subseteq \bar{R}$, with $i=1,\dots,n$, can be extended to a maximal biorder $\bar{B}_i$.

That means, for all other extensions $\bar{B}_i'$ of C_i we have

$$\bar{B}_i' \subseteq \bar{B}_i \subseteq \bar{R}.$$

The complement of $\bar{B}_i$ is then a minimal biorder B_i and $R = \bigcap_i B_i$.
This coloring procedure was successfully applied in two cases of Morse code data.
For the remaing three data sets the question whether Bidim K_{max} = Bidim R was left open after 10 hours CPU-time.

An alternative procedure turned out to be effective in all cases. Let

$\{D_i \mid i \in I\}$ be a partition of $\bar{R}$. Then $\bar{R} = \bigcup_i D_i$ and

$$R = \overline{\bigcup_i D_i} = \bigcap_i \bar{D}_i$$

If now $\{D_i\}$ is a family of biorders, each $\bar{D}_i = B_i$ is a biorder and $R = \bigcap_i B_i$. However, these biorders D_i are not necessarily maximal. Therefore, their complements B_i need not to be minimal and $|I|$ is only an upper bound of Bidim R = Chrom $H(R)$. If we denote the minimal number of colours D_i needed in this procedure by Ddim R we have

$$\text{Bidim } K_{\max} \leq \text{Bidim } R \leq \text{Ddim } R.$$

For all five data matrices of Morse codes holds

$$\text{Bidim } K_{\max} = \text{Ddim } R$$

so that Ddim R = Bidim R.
Bidim R varies between 3 and 5 for different subjects. A coloration of the empirical example shown above with Bidim R = 4 is given below.
The colors D_i are abbreviated to the indices i=1,...,4.

Example: Relation R of subject nº 4 (colored):

				A				
		a	b	c	d	e	f	
	∅							
	a	R	1	1	1	1	1	
	b	R	R	1	1	1	1	∿ ab
	c	3	3	R	2	2	1	
	d	3	3	1	R	R	1	∿ e, de
	f	2	2	2	2	2	R	
	ac	R	2	R	2	2	1	
P	ad	R	3	1	R	R	1	∿ ae, ade
	af	R	2	2	2	2	R	∿ abf
	bc	R	R	R	2	2	1	∿ abc
	bd	R	R	1	R	1	1	
	be	R	R	1	R	R	1	∿ abd, abe, bde, abde
	bf	R	R	4	2	2	R	
	cd	3	3	R	R	2	1	
	ce	3	3	R	R	R	1	∿ cde
	cf	3	3	R	2	2	R	
	df	3	3	4	R	R	R	∿ ef, def

		A						
		a	b	c	d	e	f	
	acd	R	3	R	R	2	1	
	ace	R	R	R	R	R	1	∿ bcd, bce, abcd,
	acf	R	2	R	2	2	R	abce, bcde
	adf	R	3	4	R	R	R	∿ adef
	aef	R	R	4	R	R	R	∿ bef, abef, bdef, abdef
	bcf	R	R	R	2	2	R	∿ abcf
P	bdf	4	R	4	R	R	R	
	cdf	3	3	R	R	R	R	∿ cdef
	cef	3	3	R	3	R	R	
	abdf	R	R	4	R	2	R	
	acde	R	3	R	R	R	1	
	acdf	R	R	R	R	2	R	∿ bcdf
	acef	R	3	R	R	R	R	
	bcef	R	R	R	R	R	R	∿ abcde, abcdf, abcef, acdef, bcdef, abcdef

The collection of biorders $\{B_i\}$, which is induced by the coloration process, is called a biorder representation of R. Unfortunately, in most cases the biorder representation of a binary relation is in contrast to the bidimension not unique. There may even exist a large family of representations. In order to select a special one which allows a convincing interpretation, we need a more elaborated theory about the recognition process.

6. RESUME

A lattice structure associated with the concept of lattice dimension is an appealing theoretical device to represent the empirical result of a generalized recognition experiment. This is because of its correspondence to a vectorspace. A semimodular lattice, however, imposes severe constraints on the data which are hard to be satisfied.

A multidimensional biorder representation with the concept of biorder dimension exists for each binary relation. The evaluation techniques remain to be ameliorated so that relations of larger size can be analyzed in a suitable time. In order to find a specific multidimensional biorder representation which has a rich empirical meaning, the theory of the recognition process must be developed further.

REFERENCES

[1] Beals, R., Krantz, D.H., and Tversky, A., Foundations of multidimensional scaling, Psychol. Review, 75 (1968), 127-142.

[2] Birkhoff, G., Lattice theory, Providence, American Mathematical Society (1967).

[3] Doignon, J-P., Ducamp, A., and Falmagne, J-C., On realizable biorders and the biorder dimension of a relation (to appear).

[4] Drösler, J., Grundlagen des Figur-Erkennungsexperiments, Zeitschrift für Exp. und Angew. Psychol., 27 (1980), 1-25.

[5] Herbst, K., Empirische Untersuchung eines verbandstheoretisch begründeten mehrdimensionalen Skalierungsverfahrens, Zeitschrift für Exp. und Angew. Psychol., 28 (1981), 232-254.

[6] Yannakakis, M., The complexity of the partial order dimension problem (to appear).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

ISOTONIC REGRESSION ANALYSIS AND ADDITIVITY

Ranald R. Macdonald

Department of Psychology
University of Stirling

This paper reviews the literature on isotonic regression with a view to finding tests for additivity in data whose population values are unique only up to a monotonic transformation. Several approaches are considered and a sequential one is recommended - first testing monotonicity, and then testing the double cancellation axiom. Two tests are presented for testing monotonicity and an algorithm for detecting significant departures from double cancellation is outlined.

1. INTRODUCTION

For a number of years mathematical psychologists have studied axiomatic approaches to additivity. Under the heading of conjoint measurement they have been concerned with whether the combined effects of a pair of independently manipulated variables are compatible with an underlying additive representation. To put it another way, what are the necessary and sufficient conditions under which monotonic functions of each of two variables exist such that when added together their sum is itself a monotonic function of the combined effects? A much quoted example is that of a simple bet with some probability of winning a specifed amount. Can the subjective value of such bets be expressed in terms of a monotonic function of the probability plus some monotonic function of the potential winnings? One report of a solution to the problem is given in the first paper to be published in the Journal of Mathematical Psychology, Luce and Tukey (1964), and other accounts are widely available of which the most comprehensive is Krantz, Luce, Suppes and Tversky (1971). The solution to the problem was formulated in terms of the order relations which must hold between all possible combined effects of the variables. With real data these could be found either to hold or not. Apart from a number of untestable technical axioms the data must conform to two order restrictions. They must be ordered with respect to the independent variables. That is, if one variable is held constant, an increase in the other variable should lead to an increase in the value of the data. This implies that additive data are monotonically related to the independent variables and the restriction will be referred to as the monotonicity restriction. In addition to the monotonicity restriction Luce and Tukey (1964) require that the double cancellation axiom be met. The double cancellation axiom states that if a', a'', a''' are levels of the independent variable A and b', b'', b''' are levels of the other independent variable B then if inequalities (1) and (2) hold, so should (3)

$$u_{a'',b'''} > u_{a''',b''} \quad (1)$$

$$u_{a',b''} > u_{a'',b'} \quad (2)$$

$$u_{a',b'''} > u_{a''',b'} \quad (3)$$

where u_{ab} is the value of the data corresponding to level a of A and b of B. The double cancellation axiom also is required to hold with > replaced by <. This is of limited use when dealing with the fallible data of experimentalists. In such cases it would be useful to know if any violations in the order restrictions could be explained by the inherent variability of the data or whether they represent real effects.

There is another problem in testing if the above order restrictions hold in real data. As the number of levels of the independent variables studied is finite, one cannot check the double cancellation axiom for all possible levels of A and B. In finite cases double cancellation may be satisfied while more complex cancellations are not. Any violation of a cancellation no matter how complex renders an additive representation impossible. Arbuckle and Larimer (1976) and McClelland (1977) provide probabilities of complex cancellations being violated given that monotonicity and double cancellation hold in randomly ordered matrices. These probabilities increase as the matrices increase in size. Testing all possible complex cancellations would be a hideous task. However, Luce and Tukey (1964) provide a partial way out. Where the levels of A and B are separated by equal amounts (technically A, B form a dual standard sequence) monotonicity and double cancellation are necessary and sufficient conditions to ensure an additive representation over the levels of A and B studied. Thus it is desirable to design experiments so that this condition is approximately met. If the levels of A and B do form a dual standard sequence, then it is only necessary to test monotonicity and double cancellation; in this case complex cancellations may be ignored.

It is worth distinguishing between the present conception of additivity which concerns the existence of additivity holding in some monotonic transformation of the data and additivity which holds in untransformed data. Standard tests of interactions in analysis of variance, Tukey's (1964) additivity test and Mandel's (1971) work on nonadditivity are all concerned with the second sense of additivity. The conjoint measurement problem is concerned with the first sense and if additivity exists in this sense it will also do so after any monotonic transformation has been applied. For the rest of the discussion additivity will be used only in the more general sense. Also in the applications to be considered it will be assumed that the ordering of the levels of the independent variables is known in advance.

Bartholomew (1961) presented a method for statistical inference under order restrictions which he called isotonic regression analysis. More recent reviews of the area are given in Barlow, Bartholomew, Bremner and Brunk (1972) and Smith and Macdonald (1983). A major application of isotonic regression analysis is to provide tests of differences between means subject to certain order restrictions. It would seem likely that the statistical literature on isotonic regression would be relevant to the problems in applying conjoint measurement to fallible data. This paper is designed to examine the extent to which this is true and to suggest some statistical tests for fitting an additive conjoint model.

The aims of the statisticians working on isotonic regression are somewhat different from those of the conjoint measurers. Tests developed in isotonic regression are designed to increase the power of the test when the ordering of the means can be predicted in advance. Conjoint measurement is concerned with whether there are any differences between the means incompatible with additivity rather than showing whether an additive model gives a better fit than no model at all.

2. NOTATION

The following notation applies to a two way analysis of variance, and to a one way analysis when the subscript k is omitted:

J	is the number of levels of variable A
K	is the number of levels of variable B
Y_{ijk}	is the ith observation in the jk th treatment
n	is the number of observations in each jk th treatment (assumed equal)
N	is the total number of observations
u	is the theoretical overall mean
$\hat{u}$	is the observed overall mean
u_{jk}	is the theoretical mean in the jk th treatment
$\hat{u}_{jk}$	is the observed mean in the jk th treatment
$\hat{u}_{j}$	is the observed mean in the j th level of A
$\hat{u}_{k}$	is the observed mean in the k th level of B
$\hat{\imath}u_{jk}$	is the isotonic regression estimate of u_{jk}.

$\hat{\imath}u_{jk}$ is derived with respect to some order restrictions; $u_{j'k'} < u_{j''k''}$ for some specified j', j", k', k". Barlow et al (1972) show that the isotonic regression estimate is the maximum likelihood estimate of u_{jk} consistent with these restrictions.

L is the number of different values taken by the isotonic regression estimates $\hat{\imath}u_{jk}$.

Considering first the univariate case, the usual statistic used to test for equality of treatment means is:

$$F(J-1,N-J) = \frac{\sum_{j} n(\hat{u}_j-\hat{u})^2/(J-1)}{\sum_{ij} (Y_w-\hat{u}_j)^2/(N-J)} \qquad (4)$$

In analysis of variance notation:

$$F = \frac{SSbet/dfbet}{SSw/dfw} = \frac{MSbet}{MSw} \qquad (5)$$

Bartholomew (1961) proposes the statistic E:

$$E_J^2 = \frac{\sum_{j} n(\hat{\imath}u_j-\hat{u})^2}{\sum_{ij} (Y_{ij}-\hat{u})^2} \qquad (6)$$

In our notation this is expressed:

$$E_J^2 = \frac{SSmon}{SStot} \tag{7}$$

E_J^2 differs from F in two important respects. Firstly the isotonic regression estimates are used to compute the between treatment variance through SSmon. Secondly, the between treatment variation unspecified by the order restrictions (which is referred to as SSres = SSbet - SSmon) contributes to the error term. The second difference has the affect of increasing the power of the test against $H_0(u_{j'}=u_{j''})$ when the differences between the means are in the ordering specified in $H_1(u_{j'} < u_{j''}$ when $j' < j'')$. When there are differences between the means which are neither specified nor precluded in H_1 a test using MSw as the error term may be more powerful than E_J^2.

The extension to the bivariate case in normal analyses of variance is to partition SSbet into that due to variable A (SSA), that due to variable B (SSB) and a residual called the interaction (SSAB). The degrees of freedom associated with each are known and separate F values are used to test each of the components of the between variation against the error term, see Winer (1971).

The application of isotonic regression to the two way analysis is considered by Barlow et al (1972) following Shorack (1967). They recommend using E_J^2 as defined in equation (8) rather than (7).

$$E_J^2 = \frac{SSmon(A)}{SSw + SSmon(A)} \tag{8}$$

This is to test $H_0\ u_{j'} = u_{j''}$ against $H_1\ u_{j'} < u_{j''}$ for some specified j' and j". The (A) after SSmon denotes that the order restrictions involve only levels of A. Similar tests are developed involving the B variable but no tests of differences between means in the interaction are discussed in this context.

3. UNIVARIATE TESTS OF MONOTONICITY

Let us first consider the univariate case. This is not entirely degenerate since if monotonicity were violated the data could not be represented by an additive model. A demonstration of nonmonotonicity precludes the possibility of additivity.

In univariate isotonic regression the isotonic regression estimates $\hat{i}u_j$ with respect to the order 1, 2,... J are found using the pool adjacent violators algorithm described in Barlow et al (1972). Monotonic variation in the means is then demonstrated using E_J^2 as defined in equation (6). In the application of conjoint measurement we are concerned in demonstrating nonmonotonic variation. It seems natural to take $SSres = \sum_j n(\hat{u}_j - \hat{i}u_j)^2$ as an estimate of the population variance σ^2. Nonmonotonic variation will increase the value of SSres and if it is significantly greater than would be expected from the within condition variance, the nonmonotonic variation has been demonstrated. If H_0, $u_{j'} = u_{j''}$ for all j' and j", is true then SSres is distributed as $n\sigma^2 \chi^2_{(J-L)}$. A problem arises when the null hypothesis to be tested is H_0': either $u_{j'} = u_{j''}$ or $u_{j'} < u_{j''}$, for j' < j". Suppose $u_{j'} < u_{j''}$, it is still possible that $\hat{u}_{j'} > \hat{u}_{j''}$, and as such the variation between $\hat{u}_{j'}$ and $\hat{u}_{j''}$ could contribute to SSres. The expected value of such variance would be less than that if $u_{j'} = u_{j''}$. Thus SSres is distributed as less than $n\sigma^2 \chi^2_{(J-L)}$. This argument is developed more formally by Hartigan (1967) who establishes for SSres the upper bound $n\sigma^2 \chi^2_{(J-L)}$. Thus an appropriate test statistic for H_0' might be F|L as defined in equation (9), though it will be conservative to an unknown extent for reasons given above.

$$F|L\ (J-L, N-JK) = \frac{SSres\ /\ (J-L)}{MSw} \tag{9}$$

Bartholomew (1961) considered using a F|L statistic as a test for additivity. He rejected this approach as L is a random variable which increases as the probability of H_0 decreases and H_1 increases. E_J^2 the recommended test, incorporates L into the test statistic. In the monotonic case non-limiting values of L will be determined mainly by the monotonic variation. The nonmonotonic test is concerned with testing the significance of whatever nonmonotonic variation exists independently of the size of the monotonic variation. It is not therefore appropriate to use L as part of the test statistic. Similarly it is appropriate to use SSw as the error term as it is desired to show that the nonmonotonic variation is greater than the estimate of the within condition variation. This point has been discussed above.

The conservatism of the proposed test (equation (9)) is due to the presence of small amounts of variation between means which are truly monotonic. It would seem preferable to use a test which gave greater weight to the larger nonmonotonic variation. Such a test is the studentized range test -see Winer (1971). To use this test we find the residual estimates of the treatment means ($n\hat{u}_j$) as defined in equation (10).

$$n\hat{u}_j = \hat{u}_j - i\hat{u}_j + \hat{u} \qquad (10)$$

The largest and smallest of these are substituted into the test statistic q as in equation (11).

$$q = \frac{n\hat{u}_{largest} - n\hat{u}_{smallest}}{MSw/n} \qquad (11)$$

The studentized range test is designed to detect differences between k treatment means from the difference between the largest and smallest mean. Since in the present case there are J treatments and L different isotonic estimates, the appropriate value to use for k when looking up the tables is J-L+1. For reasons analogous to those applying to F|L the range test will also be conservative but since it gives most weighting to the largest difference it should be more powerful than F|L. The relative powers of the two tests and the extent of their conservatism cannot be ascertained in general, since it will depend on the amount of true monotonic variation present in the population means.

Indeed the conservatism of these tests will depend on the true values of the population means and the probability that population means in the predicted order will give rise to sample means out of order. If the values of the isotonic estimates of the population means are treated as population values, the extent of the conservatism could be estimated by simulation. From repeated simulations the expected values of SSres and L could be found where there is no nonmonotonic variation. Corrections could be made to equation (13) to allow for this. However, the requirement to run a simulation on each test rules this out as a routine procedure.

4. BIVARIATE TESTS OF MONOTONICITY

This section attempts to generalise these results to bivariate monotonicity. The simplest approach is to rely on only the overall means for each of the levels of A and B. This is essentially the approach suggested by Shorack (1967) mentioned above, who treated the two way analysis as if it were two independent one way analyses. The isotonic regression estimates $i\hat{u}_{.j}$ are found with respect to 1, 2, ..., J and monotonic variation with respect to A tested using E_J^2 as defined in equation (8). To test the presence of nonmonotonic variation with respect to A one could therefore use F|L as defined in equation (9). The testing procedure can be repeated with respect to variable B. The two tests are mutually independent when the number of observations in each of the conditions is equal. Barlow et al (1972) discuss the case where the numbers of observations are unequal.

James (1961), commenting on Bartholomew's work, suggested ways that isotonic regression might be extended to include interactions. Bartholomew (1961) replied that he had considered the problem but found it forbidding. Hirotsu (1978) followed up these ideas by considering various comparisons of treatment means which contributed to the interaction, e.g. (12)

$$u_{j',k'} - u_{j'',k'} - u_{j',k''} + u_{j'',k''} \quad \text{for } j' < j'',\ k' < k'' \qquad (12)$$

If all such comparisons are positive (or negative) the data are monotonic. This is a stronger restriction requiring not only that the data monotonically increases with the independent variables but that the increase with each successive level of an independent variable is itself monotonically increasing (or decreasing). Another way of putting it is that the change due to the independent variables is accelerating either positively or negatively. Consequently any test which showed all such comparisons to be positive or negative would indeed establish monotonicity (though not double cancellation), but a negative result would leave open the possibility of monotonicity. Unless there is reason to expect an accelerating effect of the independent variables, this approach is probably not relevant.

Shorack's procedures outlined above are relatively straighforward but they ignore potentially relevant variation contributing to the interaction. In other words, unless the nonmonotonic variation in one variable is present averaged over all levels of the other, this approach will fail to detect the nonmonotonicity. Isotonic regression estimates iu can be found with respect to the partial order j,k < j',k' if j < j' and k < k'. Several algorithms to find isotonic regression estimates with respect to partial orders are reported in Barlow et al (1972). The pool adjacent violators algorithm which worked with simple orders fails in the more general case. As isotonic regression estimates can be found the SSbet can be split into the sums of squares between the isotonic regression estimates SSmon and SSres, a residual due to nonmonotonic variation. Again using the upper bound for SSres (Hartigan (1967)) an overall test based on $F|L$ is possible. This is presented in equation (13).

$$F|L(JK-L,N-JK) = \frac{SSres\ /\ (JK-L)}{MSw} \tag{13}$$

Similarly the multiple range test can be applied to the bivariate case. Again the residual nonmonotonic estimates of the treatment means can be found and the largest and smallest substituted into equation (11). The value for k in this case is JK-L+1.

5. EXAMPLE

The tests outlined above will be illustrated using artificial data produced by Cunningham (1982) and given in table 1 below. It will be assumed that the raw data are the means of ten independent observations (n = 10).

	Data			Isotonic Estimates		
B_3	7	6	9	6.5	6.5	9.0
B_2	3	1	8	2.67	2.67	8.0
B_1	2	4	5	2.0	2.67	5.0
	A_1	A_2	A_3	A_1	A_2	A_3

Table 1

Fictitious example from Cunningham (1982)

Applying Shorack's approach the isotonic regression estimates with respect to A (the iu's) are 3.83, 3.83, 7.33. This results in an SSmon (A) of 140 and a SSres (A) of 0.66. Nonadditivity in A would be tested by $F(1.81) = 0.66/MSw$. It would require MSw to be less than .166 to indicate significant nonadditivity at the 0.05 level. This approach would produce no evidence of nonadditivity in B, since all the variation is monotonic and SSres (B) equals zero.

Using the isotonic regression estimates with respect to the partial order $u_{j'k'} < u_{j'',k''}$ for j', j" and k', k" given in table 1, SSmon and SSres come out as 549 and 51. Nonadditivity could be tested by $F(3,81) = (51/3)/MSw$. A MSw of less than 6.83 would be required for the result to be significant at the 0.05 level. If the studentized range test were applied $q = 9.48/MSw$ and a MSw of 7.88 or less, would be required for significance at the 0.05 level.

6. DOUBLE CANCELLATION AXIOM

Monotonic data are not necessarily additive as the example due to Kruskal in Scheffe (1959) shows. Here matrix A gives the rank order of the data from smallest to largest. A is monotonic and B and C are matrices formed by a subset of the elements in A.

$$\underset{A}{\begin{bmatrix} 4 & 8 & 9 \\ 3 & 5 & 7 \\ 1 & 2 & 6 \end{bmatrix}} \qquad \underset{B}{\begin{bmatrix} 3 & 7 \\ 1 & 6 \end{bmatrix}} \qquad \underset{C}{\begin{bmatrix} 4 & 8 \\ 1 & 2 \end{bmatrix}}$$

If the data are assumed to be additive,from matrix B 1st+7th = 3rd+6th and from C 1st+8th = 4th+2nd. The left hand side of the first equation is always less than the left hand side of the second under any monotonic transformation of the data. The reverse is true of the right hand side. Clearly the additivity assumption is untenable. Data must be shown not to depart from monotonicity and not to violate double cancellation before we can conclude they are additive.

Having considered tests of monotonicity we now look at ways of testing double cancellation assuming that the data are monotonic. In its most general form we must test all the 3 x 3 matrices for inequalities (1) to (3). If we assume monotonicity many fewer matrices need to be examined. Only in 3 x 3 matrices where $a' < a'' < a'''$ and $b' < b'' < b'''$ is monotonicity compatible with inequalities (1) and (2) being true and (3) being false or vice versa. Still further reductions on the number of 3 x 3 matrices which need to be examined are possible. An algorithm for finding all the violations of double cancellation in a set of data which are compatible with monotonicity is given in the appendix.

Procedures for finding isotonic regression estimates require a priori ordering of the conditions even if it is only partial. The double cancellation axiom imposes a restriction conditional on other values of the data.

As this is not an a priori ordering the estimation algorithms given in Barlow et al (1972) would have to be extended in order for the isotonic regression estimates to be found. If the algorithms were so extended the resulting procedure for testing for non-isotonic variation would be the tests given in equations (9) and (11). In the absence of such extensions one must look for statistical tests of double cancellation using a more conventional approach.

Having tested monotonicity using one of the techniques given above and failed to find a significant violation, it is now assumed that the data are monotonic. The isotonic regression estimates with respect to bivariate monotonicity are fed into the algorithm in the appendix. This gives all the 3 x 3 submatrices from the complete data matrix which violate the double cancellation axiom. For the violation to be statistically significant each of the three inequalities contributing to the violation should be significant in its own right. Various methods exist for establishing the confidence intervals for the differences between means after an analysis of variance (see Winer (1971)). Statistically significant violations of the double cancellation axiom would have been demonstrated if the confidence intervals of the difference in all three inequalities contributing to any violation did not include zero.

Indeed, the algorithm for finding violations in double cancellation could be rewritten for looking for statistically significant violation. All this would require that the inequality signs < and > be replaced by < s and > s indicating statistically significant differences and the inequalities being treated as equal when the differences were not significant.

7. CONCLUSIONS

The overall testing procedure for additivity which seems to make most use of the data is a sequential one. The experiment should be designed so that the effects of successive levels of the independent variables are approximately equal. A test for monotonicity using either F|L (equation (13)) or the studentized range test (equation (11)) should next be performed. If this results in a significant effect monotonicity and hence additivity have been violated. If no significant effects are found, the double cancellation axiom is tested assuming monotonicity as outlined above. If neither of the above tests produces significant effects, the data have not been shown to significantly depart from additivity.

REFERENCES

[1] Arbuckle, J., and Larimer, J., The number of two way tables satisfying certain additivity axioms, Journal of Mathematical Psychology, 13 (1976), 89-100.

[2] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D., Statistical Inference under Order Restrictions, Wiley, New York (1972).

[3] Bartholomew, D.J., A test of homogeneity of means under restricted alternatives, Journal of the Royal Statistical Society, Series B, 23 (1961), 239-281.

[4] Cunningham, J.P., Multiple monotone regression, Psych. Bull., 92 (1982) 791-800.

[5] Hartigan, J.A., Distribution of the residual sums of squares in fitting inequalities, Biometrika, 54 (1967), 69-84.

[6] Hirotsu, C., Ordered alternatives for interaction effects, Biometrika, 65 (1978), 561-570.

[7] James, G.S., Contribution to discussion of the paper by D.J. Bartholomew, Journal of the Royal Statistical Society, Series B, 23 (1961), 278-279.

[8] Krantz, D.H., Luce, R.D., Suppes, P. and Tversky, A., Foundations of measurement, Academic Press, New York (1971).

[9] Luce, R.D. and Tukey, J.W., Simultaneous conjoint measurement: a new type of fundamental measurement, Journal of Math. Psychol.,1 (1964), 1-27.

[10] Mandel, J., A new analysis of variance model for non-additive data, Technometrics, 13 (1971), 1-18.

[11] McClelland, G., A Note on Arbuckle and Larimer,"The number of two-way tables satisfying certain additivity axioms", Journal of Math. Psychol., 15, (1977), 292-295.

[12] Scheffe, H., The Analysis of Variance, Wiley, New York (1959).

[13] Shorack, G.R., Testing against ordered alternatives in model 1 analysis of variance: normal theory and nonparametric, Ann. Math. Statist., 38 (1967), 1740-1753.

[14] Smith, P.T. and Macdonald, R.R., Methods for incorporating ordinal information into analysis of variance: generalizations of one-tail tests, British Journal of Mathematical and Statistical Psychology, 36 (1983), 1-21.

[15] Tukey, J.W., One degree of freedom for nonadditivity, Biometrics, 5 (1949), 232-242.

[16] Winer, B.J., Statistical Principles in Experimental Design, McGraw-Hill (1971).

APPENDIX

Algorithm for finding violations in double cancellation consistent with monotonicity

In what follows inequalities are taken to be true if and only if they hold in the strict sense. In other words if both sides of an inequality are evaluated and found to be equal the inequality will be described as both "not true" and "not false".

Let A be the complete J x K matrix containing all the data and let A^{jk} be the submatrix formed from A with columns j, j+1 and j+2 and rows k, k-1 and k-2. Let us start by considering the submatrix A^{jk} and test inequalities (1),(2) and (3) in this matrix. If and only if either (1),(2) and not (3) are true or not (1), not (2) and (3) are true is double cancellation violated. We next consider all the 3 x 3 submatrices from A with columns 1, 2, and 3, and rows K, m and n (K>m>n). The truth of (1),(2) and (3) in A^{1K} has implications for their truth in the submatrices we are considering as specified in (14) to (16) below:

$$u_{2,K} > u_{3,K-1} \Rightarrow u_{2,K} > u_{3,m} \qquad m < K-1 \tag{14}$$

$$u_{1,K-1} > u_{2,K-2} \Rightarrow u_{1,K-1} > u_{2,n} \qquad n < K-2 \tag{15}$$

$$u_{1,K} > u_{3,K-2} \Rightarrow u_{1,K} > u_{3,n} \qquad n < K-2 \tag{16}$$

With these restrictions in mind we shall consider all possible states for inequalities(1)and(3)in A.

State 1 :(1)and(3)true.

Inequalities (14) and (16) guarantee that inequalities (1) and (3) will be true for all the submatrices under consideration. Thus the double cancellation axiom is never violated.

State 2 :(1) false and (3) true.

From (16) inequality (3) will be true in all the submatrices under discussion. Because of monotonicity and inequality (3) we have:

$$u_{2,K} > u_{1,K} > u_{3,K-2} \tag{17}$$

This means that inequality (1) will be true for m = K-2 and hence because of monotonicity for all values of m not equal to K-1. Violations of double cancellation can therefore only occur when inequality (2) is false at m = K-1. We therefore test inequality (2) at m = K-1 for successively decreasing values of n starting at n = K-2 until inequality (2) is found to be not false. Monotonicity guarantees that inequality will continue not to be false for all smaller values of n. Where inequality (2) has been found to be false, violations of double cancellation have occurred.

State 3: (1) true and (3) false.

By (14) inequality (1) will be true for all the submatrices and violations will only occur where inequality (3) is false and (2) is true. The values of n for which inequality (3) is false are found by testing this inequality for successively decreasing values of n starting at K-2 until it is no longer false. Inequality (2) should now be tested for each value of n for which inequality (3) is false. This is done for all possible values of m decreasing from K-1 until inequality (2) is no longer true. Each time inequality (2) is found to be true, a violation of double cancellation has occurred.

State 4: both (1) and (3) false.

To begin with, inequality (1) is tested for decreasing values of m until at m = c' it is found to be not false and at m = c" it is found to be true: (c',c"). At this point we know we know that inequality (3) must be false at n = c'+1 (18) and since (1) is also false double cancellation will hold for n greater than c'.

$$u_{1,K} < u_{2,K} < u_{3,c'+1} \tag{18}$$

We next test inequality (3) at n = c'. If this is true inequality (1) is false for all values of m from K-1 to c'+1 and the procedures adopted in state 2 for m = K-1 should be reapplied here for all values of m from K-1 to c'+1. No violation of the type 1, 2 and not 3 is possible since 1 and 3 are false for all values of n greater than c' by monotonicity and 3 is true for all values of n greater than c'.

If c' does not equal c" inequality (1) will be equal for values of n from c"+1 to c'. Double cancellation will never be violated in these cases.

If however, inequality (3) was found to be false at n = c' no violation occurs as inequality (1) is also false for all values of m greater than c'. A violation is possible at m = c" (where inequality (1) is true) if inequality (3) is false for n = c"-1. Here the procedures adopted in state 3 should be applied.

There still remains the possibility of a violation of double cancellation if inequality (1) is false for all values of m. Here by reasoning similarly to (17) inequality (3) will be false for n greater than 1. If however, inequality (3) were found to be true at n = 1, inequality (2) should be checked from m = 2 until K-1 or inequality (2) is found to be true. Double cancellation will have been violated for every falsification of inequality (2).

State 5 : one or more of inequalities (1) and (3) are equal.

If either inequality (1) or (3) is true or both are equal the situation is like state 1. Inequalities (14) and (16) are true in the weak form (where the greater than signs are replaced by greater than or equal signs). Thus in this situation double cancellation will never be violated as neither inequality (1) nor (3) can be false.

If inequality (1) is false and inequality (3) is equal the position is similar to state 2 . The weak version of inequality (17) guarantees inequality (1) is not false for m = K-2. Violations of double cancellation can therefore only occur when m = K-1, inequality (3) is true and inequality (2) false for some value of n. Inequality (3) should be tested for m = K-1 for successively decreasing values of n until inequality (3) is found to be true at n = c' after which it will be true for all smaller values. Inequality (2) should then be evaluated at m = K-1 for values of n decreasing from n = c' until inequality (2) is no longer false. Each falsification of inequality (2) corresponds to a violation of the double cancellation axiom.

The final possibility is where inequality (1) is equal and (3) is false. Here the procedure is as in state 4 where c' and c" are different. Inequality (1) is tested for successively decreasing values of m until at m = c" it is true. Inequality (3) is evaluated at n = c"-1 and if false the procedure for state 3 is adopted.

We have now checked for double cancellation in all the 3 x 3 matrices with columns 1, 2 and 3 and rows K, m, and n. Each matrix with columns 1, o and p and rows K, K-1 and K-2 can be treated in the same way as A, for all values of $o < p < J$. This will enable the testing of double cancellation in all matrices with rows 1, o and p and columns K, m and n. This completes testing all 3 x 3 matrices including u_{1K}. In order to consider every possible violation the procedures should be repeated for all submatrices A^{jk}, $j < J-1$ and $k > 2$.

ACKNOWLEDGEMENTS

The author wishes to express his thanks to Dr. P.T. Smith for comments on the manuscript.

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

TESTING FECHNERIAN SCALABILITY BY MAXIMUM LIKELIHOOD ESTIMATION OF ORDERED BINOMIAL PARAMETERS

Rainer Mausfeld
University of Bonn

A generalization of the customary models for pair comparison data is presented in a parameter free way. For this general model, called Fechnerian structure, a maximum likelihood estimation procedure is proposed based on a two phase isotonic regression technique.

1. INTRODUCTION

There is a specific class of psychological experiments having the property that the generated data can be represented in a natural way as relative frequencies sampled from a suitably chosen probability distribution. Discrimination experiments in psychophysics and signal detectability theory, choices between gambles, preference studies and learning experiments may serve as examples. In these cases the empirical relational system is already given numerically. In its most simple form (binary response data) this system consists of a set of stimuli $(a_i)_i$ and a binary dominance relation which by independent replications leads directly to a numerical function p that satisfies: $p(a_i,a_j) + p(a_j,a_i) = 1$ for all i,j in the corresponding index set.

If with respect to this dominance relation each stimulus is assumed to be representable by a real number, and if the dominance probabilities are a monotone function of the difference of the respective stimulus parameters, the model described in the following may be appropriate (p_{ij} is shorthand for $p(a_i,a_j)$).

2. FECHNERIAN STRUCTURES

DEFINITION

Let A,B be nonempty finite sets and p be a real valued function $p:A\times B\to(0,1)$. The structure $\langle A\times B,p\rangle$ is a Fechnerian structure, if there exists a distribution function D, which is strongly monotone, whenever it is different from 0 and 1, and continuous, such that $D(0) = 1/2$ and for all $a_k, a_i \in A$ and $b_{k'}, b_j \in B$ the following condition holds:

$$p_{ij} = D\,[D^{-1}(p_{kj}) + D^{-1}(p_{ik'}) - D^{-1}(p_{kk'})] \qquad (1)$$

REMARKS

i) The definition above is independent of the numeration of elements of A or B.

ii) A Fechnerian structure is defined here by a system of constraints on the response probabilities without any explicit reference to underlying scales. These can be introduced by

$$u(a_i) = D^{-1}(p_{ik'}) - D^{-1}(p_{kk'})$$
$$v(b_j) = D^{-1}(p_{kj})$$

for any k, k'.

iii) From the injectivity of D on (0,1) it follows that for a fixed D the real valued functions u and v are unique except for a common additive constant, i.e. they are difference scales.

Thus, a structure $\langle A \times B, p\rangle$ is Fechnerian-scalable if the dominance probabilities are a monotone function of the scale value difference. For a detailed analysis of this model, see Luce-Suppes (1965) and, in a somewhat different context, Falmagne (1971). Fechnerian scalability is obviously a special case of simple scalability, a concept first introduced by Krantz (1964) and further investigated by Tversky-Russo (1969) and Tversky (1972). The observable properties of stochastic consistency and other necessary conditions on the observed relative frequencies can easily be formulated in this more general case. For instance, strong stochastic transitivity may be defined as follows.

DEFINITION

The system $\langle A \times B, p\rangle$ satisfies the condition of strong stochastic conjoint transitivity (SSCT) if for any $a,b \in A$, $x,y \in B$:

$$p(b,x) \leq \min\,[p(a,x),p(b,y)] \Rightarrow p(a,y) \geq \max\,[p(a,x),p(b,y)]\,.$$

Any Fechnerian structure satisfies SSCT. In the case of a Fechnerian utility model, (Luce-Suppes (1965)), this condition reduces to the usual strong stochastic transitivity condition.

By introducing scales into a Fechnerian structure it can be seen that it is an additive structure in the sense of conjoint measurement theory. The system of constraints on the response probabilities is in fact a system of cancellation conditions that guarantees the existence of scales in case of an additive structure. To show this consider the weak order relation $\lesssim$ on $A \times B$ that is induced by the order relation $\leq$ on the dominance probabilities:

$$(a_i,b_j) \lesssim (a_k,b_l) \Leftrightarrow p_{ij} \leq p_{kl}$$

(This definition is not quite realistic because of the artificial precision of the measurement of dominance probabilities. In order to establish inequality a minimal threshold difference that has to be exceeded is to be defined.)

It can easily be shown (Mausfeld-Stumpf (1983)) that the system $\langle A \times B, \lesssim \rangle$ satisfies an axiom scheme given by Scott (1964) for a finite additive structure, comprising all possible cancellation conditions. Thus, with respect to the qualitative laws involved, the uniqueness of scale in a Fechnerian structure is that of an ordered metric scale, whose tightness increases with the number of elements in A and B.

3. MODELS OF FECHNERIAN STRUCTURES

A Fechnerian structure is a generalization of a variety of models, including Thurstone's case V, the BTL-model, the Rasch-model, monotone pair comparison systems and strict utility models. By proper choice of the form of the response function or by particular interpretation of the sets A and B or the discrimination index η these cases may be recovered.

The set A may be interpreted as a set of stimuli, and B could be a set of persons or (as in detection problems) a set of background values. In an other context A and B may be sets of stimuli of different modalities obtained by a cross modality comparison, or even identical stimuli under different experimental conditions. In general, a particular mathematical form of the response function is assumed, the choice of which depends predominantly on mathematical convenience and its suitability for parameter estimation (cf. Thurstone (1927), p. 373; Rasch (1966), p. 3). Once a particular function is chosen, procedures for the estimation of parameters and for evaluating goodness of fit can be constructed (see e.g. Bock-Jones (1968), Andersen (1980)).

If we start with an assumption about the functional form of the response function that is more or less arbitrary as far as psychological theory is concerned, some problems may arise. It is well possible that a postulated statistical model is a powerful and parsimonious technique to summarize complex data on a descriptive level notwithstanding that it may not have any meaning in terms of substantive theory. Having obtained an acceptable goodness of fit we are faced with the question: does the invariance of scale values (that is a formal implication of the assumption of a particular mathematical function) correspond to any invariances of the hypothesised psychological structure? In other words: which is the invariance of the empirical system resulting only from purely qualitative psychological laws that are assumed to be correct?

Let us consider the following example: for the set of paired comparison data of Table 1, Guilford (1954) obtained a solution for Thurstone's case V.

Stimulus	1	2	3	4	5	6	7
1	.48	.65	.78	.92	.93	.95	.99
2	-	.50	.69	.80	.85	.97	.94
3	-	-	.48	.63	.72	.89	.91
4	-	-	-	.52	.67	.78	.86
5	-	-	-	-	.54	.64	.74
6	-	-	-	-	-	.46	.62
7	-	-	-	-	-	-	.56

Table 1

Pair comparisons matrix for seven lifted weights

The corresponding likelihood ratio test has a value of 10.18 (df=15), so Thurstone's model is not rejected. Nevertheless, as far as the underlying psychological theory is concerned, the choice of a normal response function is arbitrary. The psychological laws only refer to a Fechnerian structure.

A solution obtained in the general case by the method to be described below is shown in Table 2. The correlation between the two solutions is .998. The theoretical proportions in Table 2, however, are closer to the data in the sense of mean absolute discrepancy.

In testing Fechnerian scalability the response function is obviously not unique. If an acceptable fit is obtained under Thurstone's case V, there are infinitely many other response functions that fit the data.

Scale value	-1.13	-.83	-.24	.01	.37	.82	1.02
Stimulus	1	2	3	4	5	6	7
1	-	.6500	.7900	.8875	.9300	.9525	.9900
2	-	-	.6900	.7900	.8875	.9525	.9525
3	-	-	-	.6300	.7200	.8875	.9100
4	-	-	-	-	.6550	.7800	.8600
5	-	-	-	-	-	.6550	.7400
6	-	-	-	-	-	-	.6200
7	-	-	-	-	-	-	-

Table 2

Scale values and theoretical proportions for a Fechnerian solution for the data of Table 1.

Consequently, the proper invariance class of scale values cannot be determined by arbitrarily fixing D, but rather the whole class of appropriate response functions has to be taken into account. On the other hand, if we have obtained a bad fit, how to decide if this misfit is primarily due to violations of arbitrary mathematical assumptions or - much more important - due to violations of substantive qualitative laws that are supposed to hold?

To assume a response function that does not have an interpretation in a substantive theory "presupposes measurement as the goal and is therefore even more subject to the danger of inappropriate theorizing" (Krantz (1972)). To show that different models are equally well suited for descriptive purposes and that it is difficult to give reasons on theoretical grounds to adopt a particular model, consider a set of pair comparison data reported also by Guilford (1954) and analysed by Mosteller (1958), Hohle (1964) and Winsberg-Ramsey (1981) under varying model assumptions.

Mosteller (1958) estimated the corresponding scale values by each of the following response functions: the uniform on the interval 0 to 1, the arcsin $\sqrt{p}$, the normal, $1/2\ e^{-|x|}$, the t-distribution with 10 d.f., and he showed that the correlations between the scale values for any two methods is always greater .99 for the given data. Similar results apply to the parameters estimated under the BTL-model and under the general method to be proposed below. Even though the resulting scales are nearly linear transformations of each other (Table 4), the reproduced response probabilities under the various models are quite different (Table 5), and so is any goodness of fit test expected to be.

Vegetable	1	2	3	4	5	6	7	8	9
1. Turnips	-	.818	.770	.811	.892	.878	.899	.892	.926
2. Cabbage	-	-	.601	.723	.736	.743	.811	.845	.858
3. Beets	-	-	-	.561	.676	.736	.845	.797	.818
4. Asparagus	-	-	-	-	.588	.561	.676	.601	.730
5. Spinach	-	-	-	-	-	.507	.628	.682	.628
6. Carrots	-	-	-	-	-	-	.574	.709	.764
7. String beans	-	-	-	-	-	-	-	.527	.642
8. Peas	-	-	-	-	-	-	-	-	.628
9. Corn	-	-	-	-	-	-	-	-	-

Table 3

Preference judgements from 148 observers for each pair from nine vegetables.

	Uniform	$\arcsin\sqrt{p}$	Case V	$1/2e^{-\|x\|}$	t_{10}	BTL	Fechnerian scalability
1. Turnips	0	0	0	0	0	0	0
2. Cabbage	.284	.306	.321	.362	.327	.329	.403
3. Beets	.374	.390	.402	.433	.406	.409	.462
4. Asparagus	.578	.593	.601	.629	.605	.607	.618
5. Spinach	.683	.695	.702	.726	.705	.707	.713
6. Carrots	.669	.679	.685	.706	.688	.690	.720
7. Beans	.849	.856	.859	.871	.860	.856	.865
8. Peas	.882	.885	.886	.890	.887	.886	.880
9. Corn	1	1	1	1	1	1	1

Table 4

Scale values for Table 3 under different model assumptions.

For the proportions in Table 5 the sum of absolute deviations from the observed values is the lowest for the theoretical proportions obtained by the parameter estimation under the Fechnerian scalability model to be discussed below. The average absolute deviations in Table 5 for this model are .026.

Before presenting a method of estimating parameters in the general case of a Fechnerian structure, a short outline of the parameter estimation in the normal or logistic case might be useful.

Observed		Uniform	arcsin$\sqrt{p}$	Case V	$1/2e^{-\|x\|}$	t_{10}	ETL	Fechnerian scalability
(1,6)	.878	.863	.866	.868	.870	.869	.867	.885
(2,6)	.743	.709	.718	.724	.741	.707	.726	.737
(3,6)	.736	.660	.671	.678	.702	.681	.682	.686
(4,6)	.561	.549	.552	.554	.568	.556	.556	.583
(5,6)	.507	.492	.491	.489	.481	.488	.489	.507
(6,7)	.574	.598	.606	.611	.636	.614	.610	.607
(6,8)	.709	.616	.623	.628	.649	.630	.630	.686
(6,9)	.764	.680	.689	.696	.716	.699	.699	.696

Table 5

Comparison of observed and theoretical proportions for the "Carrot" observations (column 6 and row 6 of Table 3).

4. PARAMETER ESTIMATION IN THE CASE OF NORMAL OR LOGISTIC RESPONSE FUNCTIONS

Let N be the number of independent replications for each comparison and n_{ij} the (absolute) frequency by which a stimulus a_i dominates a stimulus a_j. Let $\bar{p}_{ij} = n_{ij}/N$. The random variable n_{ij} follows a binomial distribution with parameters N and p_{ij}, the unknown dominance probabilities to be estimated. In a complete pair comparison experiment all pairs of m stimuli are presented, leading to $\binom{m}{2}$ independent estimates $\bar{p}_{ij}$ of an equal number of binomial probabilities p_{ij}.

Let p be the vector $p = (p_{11},\dots,p_{ij},\dots,p_{m-1,m})$ and n the vector $n = (n_{11},\dots,n_{ij},\dots,n_{m-1,m})$. Then the likelihood of the entire set of independent observations is

$$L(n|p) = \prod_{i=1}^{m-1}\prod_{j=i+1}^{m}\binom{N}{n_{ij}}\, p_{ij}^{n_{ij}}\,(1-p_{ij})^{N-n_{ij}} \qquad (2)$$

Omitting additive constants the log-likelihood is:

$$\log L = \sum_{i=1}^{m-1}\sum_{j=i+1}^{m}[N.\bar{p}_{ij}.\log p_{ij} + N.(1-\bar{p}_{ij}).\log(1-p_{ij})] \qquad (3)$$

L is a function on the parameter space Θ, $\Theta \subset \mathbb{R}^n$, $n = \binom{m}{2}$. A maximum likelihood estimate $\hat{p} \in \Theta$ is to be found such that

$$L(n|\hat{p}) = \max_{p\in\Theta} L(n|p).$$

If there are no constraints on the response probabilities the ML-estimates $\hat{p}_{ij}$ obviously are the relative frequencies $\bar{p}_{ij}$. Under the assumption of a particular response function the unknown dominance probabilities in (3) are substituted by the response function, whose parameters, i.e. the scale values, are to be estimated by maximizing (3).

Following the estimation of parameters a ratio of likelihoods can be calculated: the likelihood of the data under the specific response function, that is in the restricted parameter space θ, to the likelihood under a completely general model with parameter space Θ.

Then the likelihood ratio statistic is:

$$\lambda = \frac{\max\limits_{p\in\theta} L(n|p)}{\prod\limits_{i=1}^{m-1}\prod\limits_{j=i+1}^{m}\binom{N}{n_{ij}}\bar{p}_{ij}^{\,n_{ij}}(1-\bar{p}_{ij})^{N-n_{ij}}}$$

Under conditions that are fulfilled in the normal and logistic case the distribution of -2 ln λ converges in probability to the chi-square distribution as N tends to infinity. The degrees of freedom of the limiting chi-square distribution are equal to the number of independent estimated parameters in Θ minus the corresponding number in θ. In the case discussed here, the number of independent estimated parameters in Θ is m(m-1)/2, in θ equal m-1.

For some models the values of -2 ln λ are shown in Table 6 based on the data of Table 3. According to this goodness of fit statistic the two classical models and the spline model fail to account for the empirical data. To reject Fechnerian scalability, however, it is not sufficient to reject the BTL-model or Thurstone's case V or any other model with a particular response function.

Model	-2 ln λ	df	p-value
Thurstone's Case V	52.91	28	$p < .01$
BTL-model	46.76	28	$p < .05$
Integrated B-spline- model (Winsberg-Ramsey)	42.95	25	$p < .05$
Fechnerian scalability	23.04	-	-

Table 6

5. PARAMETER ESTIMATION IN FECHNERIAN STRUCTURES

In a Fechnerian structure the only requirement on the response function D is to be continuous and strongly monotonic. Any such function guarantees the existence of an additive representation. The advantage of including all functions of theoretical interest in this definition has the drawback that the response function cannot be specified analytically in the general case. So the estimation of parameters by maximum likelihood methods is not straightforward.

In the following we will propose a method to achieve such an estimation of parameters. For sake of simplicity let now $A = B$. Assume that the rank order of the dominance probabilities with respect to the underlying scales is already known, postponing the determination of this order for a moment. Then each of these binomial parameters can be estimated by maximum likelihood estimation in the following way.

Any response function D of a Fechnerian structure gives rise to a system of inequality constraints on the response probabilities that restricts the model parameter space.

Define a vector-valued function mapping from $\mathbb{R}^n$ into $\mathbb{R}^{n-1}$ by:

$$h(p) = [h_1(p),\dots,h_{n-1}(p)]$$

with

$$\begin{aligned} h_r(p) &= h_r(p_1,\dots,p_n) \\ &= p_{r+1} - p_r \qquad 1 \le r \le n-1 \end{aligned}$$

Let $0 \le p_1 \le p_2 \dots \le p_n$ (at least one inequality strict) be an order of the response probabilities that is consistent with the model assumption of a Fechnerian structure, then the model parameter space $\theta \subset \Theta$ is:

$$\theta = \{p \mid p \in \Theta,\ h(p) > 0\}$$

Under these constraints expression (3) could numerically be maximized by any nonlinear optimization algorithm (see e.g. Gill-Murray (1974)).

Because the likelihood function is continuous in its argument it has its maximum on the closed bounded set defined by the system of inequality constraints. To show that the maximum is unique and that the ML-estimates are consistent, we rely on a correspondence between maximum likelihood estimation and isotone regression in the case considered.

THEOREM

Let $\langle A \times A, p \rangle$ be a Fechnerian structure and let $0 \leq p_1 \leq p_2 \leq \ldots \leq p_n \leq 1$ be an order on the response probabilities that satisfies the model constraints. Let $\bar{p}_t$ be the mean of a sample of N_t independent observations from a Bernoulli distribution with mean p_t. Then the maximum likelihood estimate of p_t under the prescribed order is furnished uniquely by the isotone regression of $\bar{p}_t$ with weights N_t.

This theorem is a special case of theorem 2.12. by Barlow et al. (1972), see also Ayer et al. (1955), v.Eeden (1956).

As the $\bar{p}_{ij}$ are consistent estimators of the p_{ij} in the unrestricted case, they are consistent estimators in the restricted case. For a proof, see theorem 2.2 in Barlow et al. (1972).

This result reduces algorithmic problems considerably, because the parameters can be calculated easily by means of the "pool-adjacent-violators" algorithm for isotone regression.

How to determine the linear order that is consistent with a known partial order and compatible with a Fechnerian structure.

Suppose the rank order of the stimuli is known in advance, as it generally is the case in psychophysical studies. If the stimuli are elements of a Fechnerian structure, the order on the stimuli induces a partial order on the dominance probabilities. This order has the property of simple scalability. Thus, if the rank order of the stimuli is known, the induced partial order on the dominance probabilities is given a priori. If all the other order relations, besides this partial order, could be chosen arbitrarily, a linear order could be constructed that is consistent with the partial order and whose isotone regression is identical to that one of the partial order (for the calculation of this linear order, see Barlow et al. (1972, p. 82ff)).

Any further restriction on the class of consistent linear orders will lead to a consistent linear order such that the cumulative sum diagram (CSD) is "less convex" and the slopes of the CSD are in a least squares sense less close to the slopes of the greatest convex minorant (GCM) (for details, see Barlow et al. (1972)). Thus this solution will be different from the isotone regression with respect to the partial order.

If a goodness of fit test for the isotone regression with respect to this restricted consistent linear order has proved to be satisfactory, a test for the solution with respect to the partial order will prove to be satisfactory either. On the other hand, rejecting the restricted solution leads to rejection of all solutions that are restricted either equally or more severely.

In the theory considered, the constraint to be imposed on the linear order is that it should be compatible with the assumption of a Fechnerian structure. Let us call such an order a Fechnerian order. Suppose the isotone regression has been calculated for all Fechnerian orders that are consistent with the given partial order; then choose that Fechnerian order, for which the slope of the CSD is closest to the slope of the GCM. The resulting solution is the isotone regression with respect to the partial order in the class of all Fechnerian orders, that is, it minimizes the sum

$$\sum_{i,j} [\bar{D}_{ij} - D_{ij}]^2$$

in the class of isotone functions D that satisfy condition (1).

For practical purposes, a suitable solution can be achieved by using Kruskals stress minimizing algorithm. It does not, of course, necessarily lead to the "best" Fechnerian order as defined above. But, as a conservative procedure, it can be expected to provide against wrong conclusions.

With regard to the goodness of fit statistic there remain some objections, whether the topological properties of the model parameter space - which depend on the number of stimuli - might inflate the level of significance.

If no a priori ordering of the stimuli is known in advance, as it is the case with the vegetable data, the best linear order has to be estimated entirely from the data. The use of the test statistic that will be given in the next section has no rigorous justification in this case.

Results for the Guilford-data

For the data of Table 3, the linear order that is nearest to the data in the sense of minimizing stress is determined by the UNICON-program (Roskam (1974)), with a resulting stress of .11. The ML-estimates under this order restrictions are shown in Table 7.

The mean absolute discrepancy between observed and theoretical proportions is .021, compared to .034 and .032 in the normal and logistic case respectively.

Vegetable	2	3	4	5	6	7	8	9
1. Turnips	.8075	.8075	.8345	.8850	.8850	.8955	.8955	.9260
2. Cabbage	-	.5833	.6861	.7360	.7365	.8110	.8315	.8345
3. Beets	-	-	.6066	.6861	.6861	.8075	.8075	.8315
4. Asparagus	-	-	-	.5833	.5833	.6861	.6861	.7365
5. Spinach	-	-	-	-	.5070	.6066	.6861	.6960
6. Carrots	-	-	-	-	-	.6066	.6861	.6960
7. Beans	-	-	-	-	-	-	.5270	.6066
8. Peas	-	-	-	-	-	-	-	.6066
9. Corn	-	-	-	-	-	-	-	-

Table 7

ML-estimates for a general Fechnerian structure.

It is interesting to note that the rank order of the stimuli, which is the same for all response functions given in Table 2, is reversed for stimuli 5 and 6 in the general case of a Fechnerian structure, as can be seen from the scale values in Table 2, and the comparison (5,6) in Table 3.

This result indicates a violation of the model assumptions, because in a Fechnerian structure all solutions have the invariance of an ordered metric scale.

6. EVALUATING GOODNESS OF FIT FOR FECHNERIAN STRUCTURES

Under the model hypothesis of a particular linear order on the response probabilities the parameter space unfortunately does not satisfy the conditions that are necessary for the usual asymptotic results of the likelihood ratio test statistic. The statistic $-2 \ln \lambda$ that is calculated for testing the Fechnerian model against a completely general model in §3 (Table 4) is not asymptotically χ^2-distributed and consequently is of descriptive value only. In the case of a normally distributed random variable with known variance Barlow et al. (1972) developed a test statistic for testing order hypotheses. Under the null-hypothesis of the equality of means of different experimental conditions the distribution of this statistic is a weighted sum of χ^2-distributions given by:

$$\bar{\chi}_k^2 = \sum_{i=1}^{k} n_i(\hat{\mu}_i^* - \hat{\mu})^2/\sigma^2,$$

where k is the number of means being compared, $\hat{\mu}_i^*$ is the isotone regression of the i^{th} sample mean with respect to the linear order that is tested, and $\hat{\mu}$ is the maximum likelihood estimator of the population means under the null hypothesis, that is the grand sample mean.

In order to utilize this test statistic that is appropriate for the model considered, the conditions of a normal distributed random variable with known variance are to be fulfilled. In the present case the number of independent replications is large enough to justify the approximation to the normal distribution as given by the central limit theorem.

As the variances of the proportions under H_0 depend on the unknown true dominance probabilities, the monotone transformation 2 arcsin$\sqrt{p}$ is performed on the data. It is well known that this transformation has a variance which is independent of p, namely 1/N and the transformed values are approximately distributed as a normal distribution with mean 2 arcsin$\sqrt{p}$ and variance 1/N.

If the data are in the order described by H_1, the empirical values are already the isotone regression with respect to this order and the test statistic will have its maximum. An empirical order which doesn't differ markedly from the tested order, doesn't seriously affect the power of this test (see Barlow et al. (1972, p. 165ff)). If, on the other hand, the empirical data are in such an order that the number of distinct values taken by the isotone regression tends to one, the power of the test tends to a positive value not exceeding its significance level, or to zero. Consequently, H_0 is a useful dummy hypothesis as it cannot be rejected if there are strong departures from the restrictions of the alternative hypothesis.

Results

The data of Table 3 were transformed accordingly, and the value of $\bar{\chi}_k^2$, which reduces to

$$\bar{\chi}_k^2 = N \sum_{i=1}^{k} (\hat{\mu}_i^* - \hat{\mu})^2$$

can be computed.

For the Guilford-data of Table 3 this test statistic has a value of $\bar{\chi}_{36}^2$ = 373,57. Under H_0 the probability that the test statistic takes on a value not less than this one is:

$$\Pr\{\bar{\chi}^2_{36} \geq 373{,}57\} = \sum_{j=2}^{36} P(j,36)\Pr\{\chi^2_{j-1} \geq 373{,}57\}$$

where P(j,36) is the probability that the isotonic regression function $\hat{\mu}_i^*$ takes exactly j distinct values (see Barlow et al. (1972, p. 126ff)).
By calculation of P(j,36) and by approximations of the necessary χ^2-values it can be seen that the probability of obtaining a result as large as this one is beyond any reasonable significance level.

It can safely be concluded that the stimuli can be represented as elements of a Fechnerian structure, that is, the preference scale used in these comparisons can be regarded as one-dimensional and the preference probabilities are monotone functions of the respective scale value difference.

If a Fechnerian structure cannot be accepted, the $\bar{\chi}_k^2$ test statistic provides tests for simple scalability, too. A weak test of simple scalability is to test the necessary condition that the means of the cell entries which are parallel to the matrix diagonal are in increasing order. A test that is equivalent to simple scalability is to test the complete induced partial order via isotone regression.

REFERENCES

[1] Andersen, E.B., Discrete statistical models with social science applications, Amsterdam, North Holland (1980).

[2] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., and Silverman, E., An empirical distribution function for sampling with incomplete information, The Annals of Mathematical Statistics, 26 (1955), 641-647.

[3] Barlow, R.E., Bartholomew, D.J., Bremner, J.M.,and Brunk, H.D., Statistical inference under order restrictions, London, Wiley (1972).

[4] Bock, R.D., and Jones, L.V., The measurement and prediction of judgment and choice, San Francisco, Holden-Day (1968).

[5] Eeden, C. van., Maximum likelihood estimation of ordered probabilities, Indagationes mathematicae, 18 (1956), 444-455.

[6] Falmagne, J.C., The generalized Fechner problem and discrimination, Journal of Mathematical Psychology, 8 (1971), 22-43.

[7] Gill, P.E., and Murray, W. (Eds.), Numerical methods for constraint optimization, New York, Academic Press (1974).

[8] Guilford, J.P., Psychometric methods (2 ed.), New York, McGraw-Hill (1954).

[9] Hohle, R.H., An empirical evaluation and comparison of two models for discriminability scales, Journal of Mathematical Psychology, 3 (1966), 174-183.

[10] Krantz, D.H., The scaling of small and large color differences, Unpublished doctoral dissertation, University of Pennsylvania (1964).

[11] Krantz, D.H., Measurement structures and psychological laws, Science, 175 (1972), 1427-1435.

[12] Luce, R.D.,and Suppes, P., Preference, utility and subjective probability, In R.D. Luce, R.R. Bush and E. Galanter (Eds.), Handbook of Mathematical Psychology, Vol. III, New York, Wiley (1965), 249-410.

[13] Mausfeld, R., and Stumpf, H., Latent trait-Modelle mit additiver Struktur, Zeitschrift für Differentielle und Diagnostische Psychologie, 4 (1983), 87-111.

[14] Mosteller, F., The mystery of the missing corpus, Psychometrika, 23 (1958), 279-289.

[15] Rasch, G., An informal report on the present state of a theory of objectivity in comparisons, In Proceedings of the NUFFIC International Summer session in Science at 'Het oude Hof', The Hague (1966).

[16] Roskam, E., Unidimensional conjoint measurement (UNICON) for multifaceted designs, Unpublished report, Psychological Laboratorium, University of Nijmegen (1974).

[17] Scott, D., Measurement structures and linear inequalities, Journal of Mathematical Psychology (1964), 233-247.

[18] Thurstone, L.L., Psychophysical analysis, American Journal of Psychology, 38 (1927), 368-389.

[19] Tversky, A., Choice by elimination, Journal of Mathematical Psychology, 9 (1972), 341-367.

[20] Tversky, A., and Russo, J.E., Similarity and substitutability in binary choices, Journal of Mathematical Psychology, 6 (1969), 1-12.

[21] Winsberg, S., and Ramsey, J.O., Analysis of pairwise preference data using integrated B-splines, Psychometrika, 46 (1981), 171-186.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

PROBABILISTIC CONSISTENCY, HOMOGENEOUS FAMILIES OF RELATIONS AND LINEAR Λ-RELATIONS

B. Monjardet

Université de Paris V
Centre de Mathématique Sociale

The notions of linear discrimination structure and of homogeneous family of semi-orders have been introduced in the theory of probabilistic consistency by Luce and Roberts, respectively. Here, we study the more general notions of linear Λ-relations and homogeneous families of S.T.F. relations, such notions being able to be useful in the theory of probabilistic consistency (non forced or indifference choices) and in several other situations (collective choice, non-symmetrical similarity, ...).

1. INTRODUCTION

The Roberts book, Measurement Theory (1979), contains an excellent statement of the theory of probabilistic consistency, to which we refer the reader for background and references. Here, we only recall some ideas of this theory. It deals with "pair comparison system" arising of repeated binary choice of an individual. Generally, these binary choices are "forced", so that if we denote by $p(x,y)$ the relative frequency with which x is chosen against y, we have $p(x,y)+p(y,x)=1$ (and, for convenience, it is often assumed $p(x,x)=\frac{1}{2}$). It can appear a certain "consistency" in the observed behavior of an individual (for example, $p(x,y) \geq \frac{1}{2}$ and $p(y,z) \geq \frac{1}{2}$ always imply $p(x,z) \geq \frac{1}{2}$), and the theory wants account of it. The theory has three main ingredients: explicative models, observable properties, and theorems connecting models and properties. The explicative models are based on a numerical representation of compared objects, the "utility" function, the general idea being that the more different are the utilities $u(x)$ and $u(y)$, the more easily x and y can be "discriminated", and the more different are the observed frequencies $p(x,y)$ and $p(y,x)$. For example, in the monotone scalability model

(Fishburn (1973)), there exist two maps $u : X \to \mathbb{R}$, $\Phi : u(X) \times u(X) \to [0,1]$, with Φ non decreasing (non increasing) in its first (second) argument, and $p(x,y) = \Phi[u(x),u(y)]$. One can distinguish two kinds of observable properties. Firstly, properties of the $p(x,y)$'s such as the above weak stochastic transitivity property; another example is the weak independence property (Fishburn (1973)): $p(x,z) > p(y,z)$ implies $p(x,t) \geq p(y,t)$, for all t. Secondly, properties of the "threshold relations" $R_\lambda = \{(x,y): p(x,y)>\lambda\}$, associated with the $p(x,y)$'s; for example, Roberts (1971b) defines the property: $\{R_\lambda, \lambda \in]\frac{1}{2}, 1]\}$ is an "homogeneous family" of semiorders. Finally, the theorems assert that some observable properties are necessary and sufficient ("testable") conditions for the existence of some explicative models. For instance, the weak independence, homogeneous family of semiorders and monotone scalability properties are equivalent statements (Roberts and Fishburn (1971-73)).

Now, these last results are proved for mathematical objects called "discrimination structure" (Luce (1968)), "probabilistic relation", "binary choice probability function", etc... but, in fact, these objects are special cases of the general mathematical notion of Λ-relation, i.e. a (binary) relation of which the ordered pairs are valued ("weighted") by the elements of a set Λ (some speak of "fuzzy relation"). So, we can ask two questions: can we generalize the above results to arbitrary Λ-relations? Is it useful? The first answer is positive (with Λ a linearly ordered set) and contained in theorem 13 of this paper, characterizing the "linear" Λ-relations, such a notion having been introduced by Luce (1968), see also Roubens and Vincke (1983). We now answer the second question. Firstly, in the context of probabilistic consistency, this generalization allows to tackle non forced choices ($p(x,y) + p(y,x) < 1$) and indifference choices (that can lead to $p(x,y) + p(y,x) > 1$) situations. Secondly, there are many other contexts inducing Λ-relations where the notion of linear Λ-relation can be interesting (for example, collective choice, or nonsymmetrical dissimilarity). Finally, it happens a frequent fact in mathematics: the general theory is simpler and enlightens the previous results.

In section 2, we give basic notions and results on binary relations: R-board, step-type R-board, trace T_R, "compatibility" with a weak order, Ferrers relations (Proposition 1 and Definition 2), semitransitive Ferrers relations (Propositions 3, 6 and Definitions 4 and 5). Section 3 studies the general notion of homogeneous family of relations (Definition 7 and Theorem 8), and the special case of a nested family - or chain - of relations (Proposition 9). Finally, in section 4, we define Λ-relations, Λ-chains of relations (they are one to one), and we use the results of section 3 to give eleven characterizations of linear Λ-relations (theorem 13); in conclusion, we quickly point out former results that are special cases of this theorem.

N.B. We shall use the following terms for classical binary relations: *quasiorder* = reflexive and transitive relation; *weak order* = complete quasi-order; *linear order* = antisymmetric weak order. The linear order a>b>c>d, is denoted abcd.
The dual of a relation R is the relation $R^d = \{(x,y):(y,x) \notin R\}$. The product PQ of relations P and Q is defined by x P Q y iff (if and only if) there exists z with x P z and z Q y.

2. SEMITRANSITIVE FERRERS RELATIONS

In the following sections, X always denotes a finite set, and R a binary relation on X. The notion of "board representation" of such a relation is fundamental. So, we use the following definitions.

A *R-board* is an ordered 3-uple (R,L_1,L_2) where R is a binary relation on X, and L_1 and L_2 are two linear orders on X.

A R-board is *step-type* iff

$$\text{for all } x,y,z,t \in X,\ [x\ L_1\ y,\ y\ R\ z \text{ and } z\ L_2\ t] \Rightarrow [x\ R\ t] .$$

Intuitively, a R-board corresponds to a representation of the relation R by a 0-1 tableau (Figure 1), which lines (columns) are labelled by the elements of X, ranked according to the order $L_1(L_2)$. A R-board is step-type if there exists a border line separating the 0's from the 1's on the 0-1 tableau (Figure 2).

	d	c	b	a
c	0	1	0	1
d	0	1	0	1
a	0	0	0	0
b	1	1	0	1

Figure 1: (R,cdab,dcba)

	b	d	a	c
b	0	1	1	1
d	0	0	1	1
c	0	0	1	1
a	0	0	0	0

Figure 2: (R,bdca,bdac)

If, in the R-board (R,L_1,L_2), $L_1 = L_2 = L$, we shall simply write it (R,L). In this case, the associated representation is an usual 0-1 incidence matrix.

We now define the notion of Ferrers relation.

PROPOSITION 1

Let R be a binary relation on X; the following statements are equivalent.

(1) For all $x,y,z,t \in X$, $[xRy \text{ and } zRt] \Rightarrow [xRt \text{ or } zRy]$.

(2) R has a step-type R-board, (R,L_1,L_2).

(3) There exist two maps $f,g\colon X \to \mathbb{R}$ such that for all $x,y, \in X$,
$xRy \Leftrightarrow g(x) > f(y)$.

DEFINITION 2

A *Ferrers relation* is a relation satisfying the conditions of proposition 1.

Ferrers relations have been defined by Riguet (1951) with several equivalent conditions and especially (1) and (2). They have been studied by several authors: Bouchet (1971) who gives conditions (2) and (3), Monjardet (1978), Cogis (1980,1982), Doignon, Ducamp and Falmagne (1984), who use the term biorders. In the case of a binary relation between two different sets, the corresponding notion has been used by Guttman (it is the well-known Guttman scale (1944)) and studied by Ducamp and Falmagne (1969).

It is worthwhile to notice that the "antireflexive" (or "irreflexive") Ferrers relations are the well-known (antireflexive) interval orders, whereas the reflexive Ferrers relations are the "reflexive interval orders" (i.e. the dual of interval orders).

If R is a relation on X, we set

$R^+(x) = \{y \in X\colon xRy\}$ $\quad R^-(x) = \{y \in X\colon yRx\}$
and we define the *trace* T_R of R by
$xT_Ry \Leftrightarrow R^+(x) \supseteq R^+(y)$ and $R^-(x) \subseteq R^-(y)$.
Notice that T_R is a quasiorder but not necessarily a weak order.

PROPOSITION 3 (Cogis (1980))

Let R be a relation; the following statements are equivalent.

(1) R is a Ferrers relation satisfying the following property:
for all $x,y,z,t \in X$, $[xRy \text{ and } yRz] \Rightarrow [xRt \text{ or } tRz]$.

(2) R has a step-type R-board (R,L).

(3) There exist two maps $f,g\colon X \to \mathbb{R}$, such that
for all $x,y \in X$ $xRy \Leftrightarrow g(x) > f(y)$
and $\quad f(x) \leq f(y) \Leftrightarrow g(x) \leq g(y)$.

(4) The trace T_R is a weak order.

DEFINITION 4

A *semitransitive Ferrers relation* (S.T.F) is a relation satisfying the conditions of proposition 3.

The (not very good) term "semitransitive" is due to Chipman (1960). It results of this definition that the antireflexive S.T.F. relations are the well-known (antireflexive) semiorders (A.S.O.), whereas the reflexive S.T.F relations are the reflexive semiorders (R.S.O.), i.e. the dual of semiorders. The relation R of Figures 1 and 2 is a S.T.F. relation, because the R-board (R,bdca) is step-type, but it is not a semiorder.

We shall need another characterization of a STF relation. Let R be a relation; we set

$$xPy \Leftrightarrow xRy \text{ and } yR^cx;\ xIy \Leftrightarrow xRy \text{ and } yRx$$
$$xJy \Leftrightarrow xR^cy \text{ and } yR^cx$$

DEFINITION 5

A *weak order* W is *compatible with a relation* R iff the following conditions are satisfied:

- $W \subseteq R \cup J$
- $xWyWz$ and $xIz \Rightarrow xIy$ and yIz
- $xWyWz$ and $xJz \Rightarrow xJy$ and yJz.

PROPOSITION 6

A relation R is a semitransitive Ferrers relation iff there exists a weak order compatible with R. Moreover, in this case, the set of weak orders compatible with R is the set of weak orders included in the trace T_R, and the set of linear orders L such that the R-board (R,L) is step-type is the set of linear orders included in T_R.

N.B. The above definition of compatibility and proposition 6 generalize the classical Roberts results for semiorders (1971a).

3. HOMOGENEOUS FAMILIES OF RELATIONS

We call *family of relations* on X, a n-uple $F = (R_1,\dots,R_i,\dots,R_n)$ of (binary) relations on X.

The *trace* T_F of such a family of relations is defined by

$$T_F = \bigcap_{i=1}^{n} T_i$$

where $T_i = T_{R_i}$ is the trace of the relation R_i.

The following definition generalizes the Roberts definition of an homogeneous family of semiorders (1971b).

DEFINITION 7

A *family* F *of relations* is W*-homogeneous* iff there exists a weak order W compatible with each R_i $(i=1,\dots,n)$.

THEOREM 8

Let $F = (R_1,\dots,R_i,\dots,R_n)$ be a family of relations; the following statements are equivalent.

(1) F is W-homogeneous.

(2) There exists a weak order $W \subseteq T_F$.

(3) For all i,j, $T_i \cap T_j$ is a weak order.

(4) For every i, R_i is a S.T.F. relation and for all $i \neq j$
$P_i(R_i+J_i)P_j \cup P_j(R_i+J_i)P_i \cup P_j\ P_i(R_i+J_i) \cup (R_i+J_i)P_i\ P_j \subset P_j$.

(5) The trace T_F is a weak order(compatible with each R_i).

(6) F is a W-homogeneous family of S.T.F. relations.

(7) There exists n+1 maps $f_1,\dots,f_i,\dots,f_n$, $g: X \to \mathbb{R}$ such that for every i, and for all $x,y \in X$
$xR_iy \Leftrightarrow g(x) > f_i(y)$, and $f_i(x) \leq f_i(y) \Leftrightarrow g(x) \leq g(y)$.

Moreover, if these conditions are satisfied, the weak orders W such that F is W-homogeneous are the weak orders included in T_F.

Now, we consider the special case where the family F is nested (for all i,j, $R_i \subset R_j$ or $R_j \subset R_i$). We call such a family a *chain of relations* and we write

$$F = (R_1 \supset R_2 \dots \supset R_i \dots \supset R_n).$$

In this case, we define a map n_F: $X^2 \to [0 < 1 \dots < n]$, by, for all x,y,

$$n_F(x,y) = |\{i \in [1,\dots,n] : xR_iy\}| \qquad (a)$$
$$= \operatorname{Max}\{0, \max\{i \in [1,\dots,n] : xR_iy\}\} \qquad (b)$$

Thus, n_F is a Λ-relation (see section 4), with $\Lambda = [0 < 1 < \dots < n]$.

We define the n_F-score of an element x, by:

$$s_F(x) = \sum_{z\in X} [n_F(x,z) - n_F(z,x)]$$

and the score weak order S_F by:

$$xS_Fy \Leftrightarrow s_F(x) \geq s_F(y).$$

Then, we have the following result (where we use the definitions of linearity and monotonicity of a Λ-relation given in section 4).

PROPOSITION 9

If $F = (R_1,\dots,R_i,\dots,R_n)$ is a chain of relations, the following statements are equivalent to the statements of theorem 1.

(8) $T_F = S_F$.

(9) n_F is linear.

(10) n_F has a monotone n_F-board.

Moreover, if in this case, R_n is reflexive, statement (5) becomes:

(5') For every i, R_i is a reflexive semiorder, and if we set $Q_{ij} = P_j - P_i$, for all $i < j$

$$P_i I_i Q_{ij} \cup Q_{ij} I_i P_i \cup Q_{ij} P_i I_i \cup I_i P_i Q_{ij} \subset P_j.$$

N.B. If F is an arbitrary family of relations, we can always define the relation n_F by the above formula (a). In this case, the statements of proposition 9 are necessary but not sufficient in order F is W-homogeneous.

4. APPLICATIONS TO Λ-RELATIONS: LINEAR Λ RELATIONS

Let $(\Lambda, \leqslant)$ be a linearly ordered set, with a least element 0_Λ and a greatest element 1_Λ. A Λ-*relation* ℓ on the set X is a map ℓ: $X^2 \to \Lambda$; we denote by ℓ_{xy} the image by ℓ of the ordered pair (x,y).

Following Luce (1968), we define the trace T of the Λ-relation ℓ, by:

$$xT_\ell y \Leftrightarrow \text{for every } z \in X,\ \ell_{xz} \geqslant \ell_{yz} \text{ and } \ell_{zx} \leqslant \ell_{zy}.$$

Clearly, T_ℓ is a quasiorder, not necessarily complete.

If $\Lambda \subset \mathbb{R}$, we define the ℓ-*score* of an element of (X,ℓ) by:

$$s_\ell(x) = \sum_{z \in X} (\ell_{xz} - \ell_{zx})$$

and the *score relation* S_ℓ by

$$xS_\ell y \Leftrightarrow s_\ell(x) \geqslant s_\ell(y).$$

Clearly, S_ℓ is a weak order, and $T_\ell \subseteq s_\ell$.

DEFINITION 10

A Λ-relation ℓ is *linear* iff, for all $x,y,z,t \in X$

$$[\ell_{xz} > \ell_{yz} \text{ or } \ell_{zx} < \ell_{zy}] \Rightarrow [\ell_{xt} \geqslant \ell_{yt} \text{ and } \ell_{tx} \leqslant \ell_{ty}].$$

We could define a notion of ℓ-board, generalizing the notion of R-board. But, here, we only use the following definition.

DEFINITION 11

A Λ-relation ℓ has a *monotone ℓ-board* iff, there exists a linear order L on X, such that for all $x,y,z, \in X$,

$$xLy \Rightarrow \ell_{xz} \geq \ell_{yz} \text{ and } \ell_{zx} \leq \ell_{zy}.$$

Intuitively, when ℓ has a monotone ℓ-board, it can be represented by a tableau whose the lines and the columns are labelled according to L, and with nondecreasing monotonicity of the elements in a line or in a column (see example, Figure 3).

	a	b	c	d
a	1	3	5	6
b	0	3	4	4
c	0	2	4	4
d	0	0	1	1

Figure 3

We use the following variant of the Fishburn definition of monotone scalability (1973).

DEFINITION 12

A Λ-relation ℓ is *monotonically scalable* iff there exist two maps $u: X \to \mathbb{R}$, $\Phi: u(X) \times u(X) \to \Lambda$, such that for all $x,y, \in X$, $\ell_{xy} = \Phi[u(x),u(y)]$ with Φ non increasing (non decreasing) in its first (second) argument.

Now, we show that the notion of Λ-relation is equivalent to the notion of Λ-chain of relations.

A *Λ-chain of relations* on X is a chain C of relations of X, indexed by an increasing sequence of elements of Λ: $0_\Lambda \leq \lambda_1 < \lambda_2, \ldots, < \lambda_i < \ldots \lambda_n \leq 1_\Lambda$. We write:

$$C = X^2 \supseteq R_{\lambda_1} \supset R_{\lambda_2} \ldots \supset R_{\lambda_i} \ldots \supset R_{\lambda_n} \supset \emptyset.$$

There exists a canonical bijection between the set of Λ-relations on X and the set of Λ-chains of relations on X.

- If ℓ is a Λ-relation, and if $\ell(X^2) = \lambda_1 < \lambda_2 \ldots < \lambda_n$ is the set of distinct values taken by ℓ, we define the Λ-chain C_ℓ by means of the "threshold" relations of ℓ:

$$C_\ell = X^2 \supseteq R_{\lambda_1} \ \dots \supseteq R_{\lambda_i} \quad \dots \supseteq R_{\lambda_n} \supset \emptyset$$

where $R_{\lambda_i} = \{(x,y) \in X^2 \colon \ell_{xy} \geq \lambda_i\}$.

Notice that if we set T_{λ_i} = Trace of R_{λ_i}, we have $T_\ell = \bigcap_{i=1}^{u} T_{\lambda_i}$.

- If C is a Λ-chain of relations, indexed by $\lambda_1 < \dots < \lambda_i \dots < \lambda_n$, we define a Λ-relation ℓ_C, by:

$$\ell_C(x,y) = 0_\Lambda V(V\{\lambda \in \{\lambda_1, \dots, \lambda_n\} \colon (x,y) \in R_\lambda\}).$$

N.B. The above bijection is an isomorphism between the two sets of Λ-relations and Λ-chains of relations, endowed with natural orders.

Now, we can use the section 3 results to characterize linear Λ-relations and answer a question initiated by Luce (1958).

THEOREM 13

Let ℓ be a Λ-relation on X, and C_ℓ the associated Λ-chain of relations; the following statements are equivalent.

(1) C_ℓ is W-homogeneous.

(2) There exists a weak order $W \subseteq T_\ell$.

(3) For all $\lambda_i, \lambda_j \in \ell(X^2)$, $T_{\lambda_i} \cap T_{\lambda_j}$ is a weak order.

(4) For every $\lambda_i \in \ell(X^2)$, R_{λ_i} is a S.T.F. relation, and if we set

$$Q_{\lambda_i\lambda_j} = P_{\lambda_j} - P_{\lambda_i}, \text{ for all } \lambda_i < \lambda_j$$

$$P_{\lambda_i} I_{\lambda_i} Q_{\lambda_i\lambda_j} \cup Q_{\lambda_i\lambda_j} I_{\lambda_i} P_{\lambda_i} \cup Q_{\lambda_i\lambda_j} P_{\lambda_i} I_{\lambda_i} \cup I_{\lambda_i} P_{\lambda_i} Q_{\lambda_i\lambda_j} \subset P_{\lambda_j}.$$

(5) The trace T_ℓ is a weak order (compatible with each R_{λ_i}).

(6) C_ℓ is a W-homogeneous family of S.T.F. relations.

(7) There exists n+1 maps $f_1, \dots, f_i, \dots, f_n$, g such that for every i and for all $x,y \in X$

$$xR_{\lambda_i}y \Leftrightarrow g_i(x) > f(y), \text{ and } f_i(x) \leq f_i(y) \Leftrightarrow g(x) \leq g(y).$$

(8) $T_\ell = S_\ell$.

(9) ℓ is linear.

(10) ℓ has a monotone ℓ-board.

(11) ℓ is monotonically scalable.

Notice that condition (8) gives a very efficient way to recognize if a Λ-relation ℓ is linear: for all x,y with $xS_\ell y$, to test if $xT_\ell y$.

Now, we consider particular cases of linear Λ-relations.

(1) If for every x, $\ell_{xx} = \alpha$, then for every $\lambda \geq \alpha (\lambda < \alpha)$, R_λ is a reflexive (antireflexive) semiorder.

(2) If Λ is a set of numbers, and if for all x,y, $\ell_{xy} + \ell_{yx} = \beta$, we say that ℓ is ipsodual; then, the threshold relations R_λ, for $\lambda < \frac{\beta}{2}$, are the dual of the R_λ, for $\lambda \geq \frac{\beta}{2}$. Thus, we can only consider the "half" of the chain C_ℓ of relations. The results obtained in this case by Jacquet-Lagrèze (1978) and Roubens and Vinkce (1983), correspond respectively to equivalences of statements (6), (8), (10) and (7), (9), (10), of theorem 13.

(3) If for all x,y, $\ell_{xy} + \ell_{yx} = \beta$, and moreover $\ell_{xy} = \frac{\beta}{2}$ implies $xT_\ell y$ and $yT_\ell x$, one obtains results of Roberts on homogeneous families of semiorders (1971b).

(4) Finally, in the case of an homogeneous family of two semiorders $(P \cup Q, P)$, condition (4) becomes $P\,I\,Q \cup Q\,I\,P \cup Q\,P\,I \cup I\,P\,Q \subset P$. This case has been studied by Cozzens and Roberts (1982), and Roy and Vincke (1984); for example, they both give necessary and sufficient representations in order that numerical representations with two constant thresholds exist (i.e. there exist f_i,g with

$$xPy \text{ iff } g(x) > f_2(y),\ x(P\cup Q)y \text{ iff } g(x) > f_1(y), \text{ and } f_2-g \equiv \lambda > f_1-g \equiv \mu > 0).$$

REFERENCES

[1] Bouchet, A., Etude combinatoire, applications, Thèse, Université Scientifique et Médicale de Grenoble (1971).

[2] Chipman, J.S., Consumption theory without transitive indifference, in Preferences, Utility and Demand, New York, Harcourt Brace (1971).

[3] Cogis, O., Détermination d'un préordre total contenant un préordre et contenu dans une relation de Ferrers lorsque leur domaine commun est fini, C.R.A.S., 283 Série A (1976), 1007-1009.

[4] Cogis, O., A propos des quasi-ordres - Note, Math. Sci. Hum., 72 (1980) 107-111.

[5] Cogis, O., Ferrers digraphs and threshold graphs, Discr. Math. 38 (1982), 33-46.

[6] Cozzens, M.B., and Roberts, F.S., Double semiorders and double indifference Graphs, SIAM, J. Alg. Discr. Math., 3-4 (1982), 566-583.

[7] Doignon, J.-P., Ducamp, A. and Falmagne, J.C., On realizable biorders and the biorder dimension of a relation, J. Math. Psych., to appear (1984).

[8] Ducamp, A., and Falmagne, J.C., Composite measurement, J. Math. Psych., 6 (1969), 359-390.

[9] Fishburn, P.C., Binary choice probabilities: on the variety of stochastic transitivity, J. Math. Psych., 10-4 (1973), 327-352.

[10] Jacquet-Lagrèze, E., Représentations de quasi-ordres et de relations probabilistes sous forme standard et méthodes d'approximation, Math. Sci. Hum., 63 (1978), 5-24.

[11] Luce, R.C., A probabilistic theory of utility, Econometrica, 26 (1958), 193-224.

[12] Mirkin, B.G., Description of some relations on the set of real-line intervals, J. Math. Psych., 9 (1972), 243-252.

[13] Monjardet, B., and Jacquet-Lagrèze, E., Modélisation des préférences et quasi-ordres, Math. Sci. Hum., 62 (1978), 5-10.

[14] Monjardet, B., Axiomatiques et propriétés des quasi-ordres, Math. Sci. Hum., 63 (1978), 51-82.

[15] Riguet, J., Les relations de Ferrers, C.R.A.S., 232, Paris (1951), 1729-1730.

[16] Roberts, F.S., On the compatibility between a graph and a simple order, J. Comb. Th., 11 (1971a), 248-263.

[17] Roberts, F.S., Homogeneous families of semiorders and the theory of probabilistic consistency, J. Math. Psych., 8 (1971b), 248-263.

[18] Roberts, F.S., Measurement theory, Encyclopaedia of mathematics and its applications, Reading, Mass., Addison-Wesley (1979).

[19] Roubens, M., and Vincke, Ph., Linear fuzzy graphs, Fuzzy sets and systems, 10 (1983), 79-86.

[20] Roy, B., and Vincke, Ph., Pseudo-orders: definition, properties and numerical representation (1983), submitted.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

APPLICATIONS OF THE THEORY OF MEANINGFULNESS TO ORDER AND MATCHING EXPERIMENTS

Fred S. Roberts

Department of Mathematics
Rutgers University
New Brunswick, New Jersey, USA

The theory of meaningfulness of statements using numerical scales is applied to the analysis of statements involving scales representing the semiorders and indifference graphs which result from order and matching experiments.

1. INTRODUCTION

The concept of meaningfulness is among the most useful concepts of measurement theory. This concept has found applications to such areas as measurement of average performance, importance ratings, consumer price indices, pollution measures, indices of consumer confidence, psychophysical scaling, analysis of social networks, structural modeling in decisionmaking, and the dimensional invariance of scientific laws. Such applications are surveyed in Roberts (1979, 1983a). In this paper, we apply the theory of meaningfulness to the analysis of order and matching experiments when matching is not transitive.

The paper is organized as follows. Section 2 describes what we mean by order and matching experiments. Section 3 summarizes the relevant concepts from the theory of meaningfulness in measurement. Finally, Section 4 contains the main results and Section 5 gives the proofs.

2. ORDER AND MATCHING EXPERIMENTS

A variety of experiments involve judgements such as "I am indifferent between x and y", "x and y are similar", "x and y are the same", "I cannot tell x and y apart", or "x and y match". We shall call these matching experiments. Related experiments involve judgements of order, such as "I prefer x to y", "x is bigger than y", "x is better than y", and so on. We shall call these order experiments. Coombs (1964) distinguishes these two kinds of experiments by talking about proximity relations and order relations. For discussion of order and matching experiments, see, for instance, Galanter (1956), Goodman (1951), or Restle (1959).

It was observed by Armstrong (1939) that observed matching is often not transitive. Similar observations go back to Poincaré in the nineteenth century. In response to such observations, Luce (1956) introduced the notion of semiorder to account for order (preference) judgements when matching (indifference) is not transitive. In particular, if A is a finite set and P is a binary relation on A, (A,P) is a semiorder if we can find a function f: A → Re satisfying

$$aPb \Leftrightarrow f(a) > f(b) + \delta, \qquad (1)$$

where δ is a fixed positive number measuring threshold or just noticeable difference. Semiorders were characterized by Luce (1956) and by Scott and Suppes (1958). Analogous to semiorders are indifference graphs. If A is a finite set and I is a binary relation on A, (A,I) is an indifference graph if we can find a function f: A → Re satisfying

$$aIb \Leftrightarrow |f(a)-f(b)| \leq \delta, \qquad (2)$$

where again δ is a fixed positive number measuring threshold or just noticeable difference. Indifference graphs were characterized by Roberts (1969).

In this paper, we shall study the meaningfulness of various assertions involving the functions f which satisfy Equations (1) and (2).

3. MEANINGFULNESS

The theory of meaningfulness of statements involving scales of measurement stems from the work of Suppes (1959) and Suppes and Zinnes (1963), though it has its roots in the notion of invariance which goes back much earlier. The theory is summarized in Roberts (1979). We give only the most basic definitions here, referring the reader to Roberts (1979) for all undefined terms.

Suppose $\mathcal{A}$ and $\mathcal{B}$ are two relational systems, with A the underlying set of $\mathcal{A}$ and B the underlying set of $\mathcal{B}$. If f: A → B is a homomorphism from $\mathcal{A}$

into $\mathcal{B}$, we call the triple $(A,\mathcal{B},f)$, and sometimes just the function f, a scale. It is a numerical scale if B is a set of real numbers. A function $\phi: f(A) \to B$ is called an admissible transformation of the scale f if $\phi \circ f$ is a scale. Suppes (1959) and Suppes and Zinnes (1963) say that a statement involving numerical scales is meaningful if its truth value is unchanged if admissible transformations are applied to every scale in the statement. Unfortunately, as Roberts and Franke (1976) point out, there are some difficulties with this definition. To see this, it suffices to consider the semiorder defined by

$$A = \{a,b,c\},\ P = \{(c,a),(c,b)\}. \tag{3}$$

Then two homomorphisms f and g satisfying Equation (1) with $\delta = 1$ are given by

$$f(a) = 0,\ f(b) = 0,\ f(c) = 2 \tag{4}$$

and

$$g(a) = .1,\ g(b) = 0,\ g(c) = 2. \tag{5}$$

It follows that

$$f(a) = f(b), \tag{6}$$

and moreover that

$$(\phi \circ f)(a) = (\phi \circ f)(b)$$

for any admissible transformation ϕ of f. Hence, by the Suppes-Zinnes definition, the statement (6) is meaningful. However, we also have

$$g(a) > g(b). \tag{7}$$

The assertion that a and b receive the same scale values is satisfied for one homomorphism and violated for another. It seems meaningless. This leads us to modify the definition as follows. We say a statement involving numerical scales is meaningful if its truth value is unchanged if every scale in the statement is replaced by another acceptable one. Under this more general definition, statement (6) is meaningless. This definition is widely accepted, at least in spirit. For alternative formulations, see Adams, et al (1965), Falmagne and Narens (1983), Luce (1978), Narens (1981), Pfanzagl (1968), Roberts (1980), Roberts and Franke (1976), and Robinson (1963).

A scale $(A,\mathfrak{B},f)$ is called regular if for any other scale $(A,\mathfrak{B},g)$, there is a function ϕ on $f(A)$ so that $g = \phi \circ f$. It is exactly the irregularity of the homomorphism f of Equation (4) which causes difficulties. There is no ϕ on $f(A)$ so that $g = \phi \circ f$, for g as defined in (5). Roberts and Franke (1976) observe that if every scale in question is regular, then the original Suppes-Zinnes definition of meaningfulness can be used, but otherwise the more general definition is necessary.

4. THE ANALYSIS OF ORDER AND MATCHING EXPERIMENTS

In this section, we apply the concept of meaningfulness to statements involving scales satisfying Equations (1) and (2). Perhaps the most common kind of assertion using scales of measurement is the assertion

$$f(a) > f(b). \tag{8}$$

See Roberts (1983b) for a general treatment of the meaningfulness of the assertion (8). We shall investigate this and related assertions here.

For the indifference graph representation (2), the conclusion (8) is in general meaningless. For if f satisfies (2), then so does $-f$. The scales f and g in Equations (4) and (5) show that a conclusion (8) can also be meaningless for the semiorder representation (1). For $f(a) > f(b)$ fails, but $g(a) > g(b)$ holds.

Suppose next that we define a semiorder by taking

$$A = \{a,b,c\}, \quad P = \{(a,b),(b,c),(a,c)\}.$$

Then for every f satisfying (1), we have $f(a) > f(b) + \delta$, so the conclusion $f(a) > f(b)$ is meaningful. In fact, in this example, $f(x) > f(y)$ is meaningful for every $(x,y) \in A \times A$. We shall see under what conditions a similar conclusion can be drawn.

We say that xEy holds if for all z, $xPz \Leftrightarrow yPz$ and $zPx \Leftrightarrow zPy$. It is clear that (A,E) is an equivalence relation. In the example (3), aEb, we have the following result:

Theorem 1: Suppose A is finite and f is a homomorphism on (A,P) satisfying (1). Then $f(x) > f(y)$ is meaningful for all $(x,y) \in A \times A$ if and only if whenever $x \neq y$, $\sim xEy$.

For a proof of Theorem 1, see the next section.

Of course, we could not expect the conclusion $f(a) > f(b)$ to be meaningful for a matching experiment, i.e., one where we just have matching but no order judgements. On the other hand, the notion of order is often captured by judgments of betweenness, and we might expect judgments of betweenness to be meaningful. We say that a real number β is between two real numbers α and γ if $\alpha < \beta < \gamma$ or $\gamma < \beta < \alpha$. If f satisfies (2), is it meaningful to conclude that $f(b)$ is between $f(a)$ and $f(c)$?

To answer this question, let us say xEy holds if for all z, $xIz \Leftrightarrow yIz$. Again, it follows that (A,E) is an equivalence relation.
For instance, suppose

$$A = \{a,b,u,v,\alpha,\beta\}$$

$$\begin{aligned} I = \{&(a,b),(b,a),(u,\alpha),(\alpha,u),(u,\beta),(\beta,u),(v,\alpha),(\alpha,v),(v,\beta),(\beta,v),\\ &(\alpha,\beta),(\beta,\alpha)\} \\ &\cup \{(x,x)\colon x \in A\}. \end{aligned} \tag{9}$$

Then we have $\alpha E\beta$. We say that (A,I) is connected if for all x and y in A, there are $x_1,x_2,\ldots,x_n$ so that $xIx_1, x_1Ix_2, \ldots, x_{n-1}Ix_n, x_nIy$. (A,I) of (9) is not connected, as the above fails with $x = a$, $y = u$. We have the following result:

Theorem 2: Suppose A is finite and f is a homomorphism on (A,I) satisfying (2), and suppose that $|f(A)| \geq 3$. Then the conclusion that $f(y)$ is between $f(x)$ and $f(z)$ is meaningful for all x,y,z in A if and only if the following conditions hold:

(a) For all $x \neq y$, $\sim xEy$.

(b) (A,I) is connected.

For a proof of Theorem 2, see the next section. To illustrate this theorem, let us take $\delta = 1$ and define f and g for (A,I) of (9) as follows:

$$f(a) = 0,\ f(b) = .6,\ f(u) = 4,\ f(v) = 5.1,\ f(\alpha) = 4.2,\ f(\beta) = 4.3,$$

and

$$g(a) = 10,\ g(b) = 9.4,\ g(u) = 3,\ g(v) = 4.1,\ g(\alpha) = 3.3,\ g(\beta) = 3.2.$$

Then we have $f(\alpha)$ between $f(a)$ and $f(v)$, but $g(\alpha)$ not between $g(a)$ and $g(v)$; we also have $f(\alpha)$ between $f(u)$ and $f(\beta)$, but $g(\alpha)$ not between $g(u)$ and $g(\beta)$. Both conditions of Theorem 2 are violated: we have $\alpha E \beta$ and we have (A,I) disconnected.

To summarize: Betweenness conclusions can be drawn from matching experiments, but only under certain conditions.

5. PROOFS OF THEOREMS 1 AND 2

To prove Theorem 1, we use the notion of simple order (A,R) compatible with the binary relation (A,P) which is introduced in Roberts (1971). Suppose (A,R) is a simple order and (A,P) is an asymmetric binary relation. Define I on A by

$$xIy \Leftrightarrow \sim xPy\ \&\ \sim yPx.$$

We say (A,R) is compatible with (A,P) if

$$xPy \Rightarrow xRy \tag{10}$$

and

$$xRyRz\ \&\ xIz \Rightarrow xIy\ \&\ yIz. \tag{11}$$

Assuming A is finite, Roberts (1971) proves the following:

[1] If (A,P) is an asymmetric binary relation, then there is a simple order R on A compatible with (A,P) if and only if (A,P) is a semiorder.

[2] Every compatible simple order R on A is obtained as follows: there is

a unique simple ordering R^* of equivalence classes under E, and we then obtain R from R^* by ordering elements within an equivalence class arbitrarily.

[3] If R is a compatible simple order on A, then there is a homomorphism f (function satisfying (1)) so that

$$xRy \Leftrightarrow f(x) > f(y). \tag{12}$$

Conversely, if f is a homomorphism and R is defined by (12), and $\sim xEy$ for all $x \neq y$, then R is a compatible simple order on A.

Proof of Theorem 1. Suppose xEy, $x \neq y$. By [2] above, there is a simple order R on A, compatible with P, in which x follows y, and another R' in which y follows x. By [3] above, there are homomorphisms f and g so that $f(x) > f(y)$ and $g(x) < g(y)$. We conclude that $f(x) > f(y)$ is meaningless. Conversely, suppose $x \neq y$ implies $\sim xEy$. Then by [2], there is a unique compatible simple order R. Moreover, by [3], every homomorphism f is related to R by (12). Thus, for every $(x,y) \in A\times A$, and for all homomorphisms f and g,

$$f(x) > f(y) \text{ iff } g(x) > g(y).$$

Q.E.D.

To prove Theorem 2, suppose (A,I) is reflexive and symmetric. Let us say that a simple order R on A is compatible with I if Equation (11) holds. Then assuming A is finite, Roberts (1971) proves the following:

[4] If (A,I) is a reflexive, symmetric binary relation, then there is a simple order on (A,I) compatible with (A,I) if and only if (A,I) is an indifference graph.

[5] Suppose that (A,I) is connected. Then every compatible simple order R on A is obtained as in [2] above, except that for each such simple order, reversing the entire order gives another compatible simple order.

[6] If (A,I) is not connected, its connected components are its maximal induced connected subrelations. We find all compatible simple orders R on A as follows. Find one compatible simple order for each connected component of (A,I), and then order these orders arbitrarily.

[7] Statement [3] above holds (with homomorphisms being functions satisfying (2)).

Proof of Theorem 2. Suppose $x \neq y$ and xEy. Since $|f(A)| \geqslant 3$, we can find $z \neq x,y$. Now by [5] and [6], there are compatible simple orders R, in which x follows y follows z, and R', in which y follows x follows z. Thus, by [7], there are homomorphisms f and g with $f(x) > f(y) > f(z)$ and $g(y) > g(x) > g(z)$. The conclusion that f(y) is between f(x) and f(z) is meaningless.

Suppose $G = (A,I)$ is not connected. Thus, there are at least two connected components. If there are three connected components, K_1, K_2, K_3, these components can be ordered arbitrarily in a compatible simple order (by [6]). Fix x in K_1, y in K_2, z in K_3. Then y is between x and z in some compatible simple order and not in another. Then by [7], f(y) is between f(x) and f(z) in some homomorphism, but g(y) is not between g(x) and g(z) in another homomorphism. Next, suppose there are two connected components, K_1 and K_2. Since $|f(A)| \geqslant 3$, at least one of these components, say K_1, has two elements x and y. Let z be in K_2. Then by [5], there are compatible simple orders of K_1 in which x follows y and in which y follows x. If we order K_2 before K_1, we obtain compatible simple orders for (A,I) in which x follows y follows z and in which y follows x follows z. As above, we can find homomorphisms f and g so that f(y) is between f(x) and f(z) and g(y) is not between g(x) and g(z).

Finally, suppose (a) and (b) hold. Then [5] says that there are exactly two compatible simple orders on (A,I), one the reverse of the other. Hence, y is between x and z in both of these, or in neither. Thus by [7], f(y) is between f(x) and f(z) in all homomorphisms or in none.

Q.E.D.

REFERENCES

[1] Adams, E.W., Fagot, R.F., and Robinson, R.E., A Theory of Appropriate Statistics, Psychometrika, 30 (1965), 99-127.

[2] Armstrong, W.E., The Determinateness of the Utility Function, Econ. Jrnl., 49 (1939), 456-467.

[3] Coombs, C.H., A Theory of Data, Wiley, New York, (1964).

[4] Falmagne, J.C., and Narens, L., Scales and Meaningfulness of Quantitative Laws, Synthese, 55 (1983), 287-325.

[5] Galanter, E.H., An Axiomatic and Experimental Study of Sensory Order and Measure, Psychol. Rev., 63 (1956), 16-28.

[6] Goodman, N., Structure of Appearance, Harvard University Press, Cambridge, (1951).

[7] Luce, R.D., Semiorders and a Theory of Utility Discrimination, Econometrica, 24 (1956), 178-191.

[8] Luce, R.D., Dimensionally Invariant Numerical Laws Correspond to Meaningful Qualitative Relations, Philos. Sci., 45 (1978), 1-16.

[9] Narens, L., A General Theory of Ratio Scalability With Remarks About the Measurement-theoretic Concept of Meaningfulness, Theory and Decision, 13 (1981), 1-70.

[10] Pfanzagl, J., Theory of Measurement, Wiley, New York, (1968).

[11] Restle, F., A Metric and an Ordering on Sets, Psychometrika, 24 (1959), 207-220.

[12] Roberts, F.S., Indifference Graphs, in F. Harary (ed.), Proof Techniques in Graph Theory, Academic Press, New York (1969), 139-146.

[13] Roberts, F.S., On the Compatibility between a Graph and a Simple Order, J. Comb. Theory, 11 (1971), 28-38.

[14] Roberts, F.S., Measurement Theory, with Applications to Decisionmaking, Utility, and the Social Sciences, Addison-Wesley, Reading, MA (1979).

[15] Roberts, F.S., On Luce's Theory of Meaningfulness, Philos. Sci., 47 (1980), 424-433.

[16] Roberts, F.S., Applications of the Theory of Meaningfulness to Psychology, mimeographed, Department of Mathematics, Rutgers University, New Brunswick, NJ, (1983a).

[17] Roberts, F.S., On the Theory of Meaningfulness of Ordinal Comparisons, Measurement, (1983b), to appear.

[18] Roberts, F.S. and Franke, C.H., On the Theory of Uniqueness in Measurement, J. Math. Psychol., 14 (1976), 211-218.

[19] Robinson, R.E., A Set-theoretical Approach to Empirical Meaningfulness of Empirical Statements, Tech. Rept. No. 55, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, CA, (1963).

[20] Scott, D., and Suppes, P., Foundational Aspects of Theories of Measurement, J. Symbolic Logic, 23 (1958), 113-128.

[21] Suppes, P., Measurement, Empirical Meaningfulness and Three-Valued Logic, in C.W. Churchman and P. Ratoosh (eds.), Measurement: Definitions and Theories, Wiley, New York (1959), 129-143.

[22] Suppes, P., and Zinnes, J., Basic Measurement Theory, in R.D. Luce, R.R. Bush, and E. Galanter (eds.), Handbook of Mathematical Psychology, Vol. 1, Wiley, New York (1963), 1-76.

ACKNOWLEDGEMENTS

The author would like to thank the National Science Foundation for its support under Grant Number IST-83-01496 to Rutgers University. He would also like to thank Arundhati Ray-Chaudhuri, Larry Harvey, Sam Rosenbaum, and Jeff Steif for their helpful comments.

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

A NEW DERIVATION OF THE RASCH MODEL

Edw. E. Roskam
Department of Psychology, University of Nijmegen
and
Paul G.W. Jansen
Rijks Psychologische Dienst, Den Haag

The Rasch model is usually derived from statistical requirements of estimation of its parameters. The principle of 'specific objectivity' or, equivalently, 'sample independence' requires that subject parameters can be estimated independent from the item parameters and vice versa. This in turn requires that sufficient statistics exist for the parameters, from which the Rasch model follows. Rather than using a statistical argument, the present paper presents a derivation of the Rasch model based on the requirement that the probability of inferential ordering of subjects (or items, resp.) is sample independent. The derivation parallels Ducamp & Falmagne's axiomatization of the Guttman scale (composite measurement); it is first given for the ordering of two items or subjects, and subsequently generalized to the probability of a multiple ordering.
The core equation in the derivation is, of course, not different from the core equation in the statistical derivation.

1. INTRODUCTION

As a motivation for using the Rasch model, the measurement argument of specific objectivity is not separated very well from the statistical argument of sufficiency. The measurement argument stipulates that subject measurement should not be influenced by the selection of a specific measuring agent (item) from the reference universe, whereas the statistical argument stresses the statistical advantages of using conditional estimation procedures (e.g. consistent estimators and test statistics with known and desirable properties). In particular, this requires the existence of sufficient statistics for the parameters of the Rasch model.

There is a need for an exact definition of specific objectivity in a probabilistic context whose formulation is independent of purely statistical considerations. A recently proposed definition (Roskam (1983)) of a specifically objective probabilistic comparison is discussed and elaborated in section 2. In section 3 we argue subsequently why we consider the principle of specific objectivity to be such an important requirement of scientific comparisons, including the ones of attitude theory.

RASCH'S FORMULATION OF SPECIFIC OBJECTIVITY

Rasch (1966c, 4-5) introduced specific objectivity in terms of a theory about objectivity in deterministic comparisons between objects by means of an agent.

An example of a specifically objective comparison is the relation between mass m, force F and acceleration a, in Newton's Second Main Law:

$$F = ma \qquad (1)$$

Thus: the force F_i that acts upon a body v with mass m_v causes an acceleration

$$a_{vi} = \frac{F_i}{m_v} \qquad (2)$$

which can be observed. Now suppose we want to compare two objects v and w with respect to their masses m_v and m_w respectively. To this end, we let a force, F_i say, act upon these objects, and observe the reactions a_{vi} and a_{wi}. The ratio comparison

$$\frac{a_{vi}}{a_{wi}} \qquad (3)$$

is, due to (2), independent of the specific force F_i that was used to bring about the reactions a_{vi} and a_{wi}:

$$\frac{a_{vi}}{a_{wi}} = \frac{F_i/m_v}{F_i/m_w} = \frac{m_w}{m_v} \qquad (4)$$

Thus the comparison of the objects is independent of the agent i: it is specifically objective. This example is due to Rasch (1960, ch. VII). In his treatment he stresses the multiplicative conjunction of F_i and m_v in (2), which is understandable considering (4).

Then he states (1.c):

> I wish to point out that in this context only the objects and/or the agents are subject to comparison, while the data themselves are not directly compared; they only serve as instruments for the comparison aimed at. The consequences of these two concepts: (specific) comparisons and specific objectivity, completed by the requirement that a comparison is always possible and the result always unambiguous, are really overwhelming.

This "unambiguity" may be clarified by means of the following quotation from Rasch (1960, 9):

> If a relationship between two or more statistical variables is to be considered really important, as more than an ad hoc description of a very limited set of data (...), the relationship should be found in several sets of data which differ materially in some relevant respects.

In general, the comparison of objects v and w from a universe of objects is said to be specifically objective when the result of this comparison is the same for all agents i,...,K from universe of agents that are used to elicit reactions $\lambda(v,i)$ from these objects. So analogous to the relation between mass, force and acceleration, we have:

$$\lambda_{vi} = \theta_v * \varepsilon_i \tag{5}$$

and a necessary and sufficient condition for specifically objective measurement of either objects or agents. Thus, in a deterministic context, $\lambda(v,i)$ should be multiplicatively or additively decomposable to allow for specifically objective comparisons (cf. Rasch (1979), 79).

Ramsay (1975), in his review of Krantz et al. (1971), states that this book

> "deals with the fact that virtually all the laws of physics can be expressed numerically as multiplications or divisions of measurements" (o.c., 258).

And on p. 262 he utters again his amazement at

> "the remarkable fact that throughout the gigantic range of physical knowledge numerical laws assume a remarkable simple form provided fundamental measurement has taken place (...)".

Thus it seems that specific objectivity is a property of 'virtually all' comparisons in physics.

Above, specific objectivity has been defined in a deterministic context. In a probabilistic model, the parameters λ, θ and ε are essentially unobservable. Therefore, in the literature the requirement for specific objectivity is reframed as the demand that the measurement model should allow for "probability expressions which depend on only one set of parameters at a time" (Douglas (1980), 1), and consequently specific objectivity is defined as follows for probabilistic models:

> "A two parameter statistical model for measuring is defined to be specifically objective if statistical inference procedures exist such that if we intend to draw statistical inference about only one of the parameters it does not depend on other parameters" (Scheiblechner, (1979), 21).

and

> "(...) the conditional distribution given sufficient statistics can be taken as a specifically objective statistical model" (o.c., 22).

A similar definition is given by Fischer & Pendl ((1980), 173-174). In a probabilistic context, specific objectivity is also called 'sample independence' or 'sample-free measurement'. By these definitions specific objectivity in a probabilistic context becomes equivalent to the statistical property of sufficiency, so that, in fact, specific objectivity coincides with the Rasch model. Suppes (in Spada & Kempf (1977), 22) questioned the possibility of finding a general formulation for specific objectivity that could be separated from the Rasch model itself. In a reply Fischer (o.c., 36) admitted that "it is difficult to give a completely general definition of specific objectivity".

Thus it seems that the measurement argument of specific objectivity and the statistical argument of sufficiency are not well separated in the literature of the Rasch model.

2. A NEW FORMULATION OF SPECIFIC OBJECTIVITY IN PROBABILISTIC COMPARISONS

Roskam (1981, 57; 1983, 81-83) demonstrated that the Rasch model is equivalent to the requirement that the probability of inferring the item ordering iRj is the same for all subjects from the reference population. In agreement with the first quotation of Rasch, two items i and j ("objects") are ordered ("compared") on account of the (dichotomous) responses a_{vi} and a_{vj} of a subset v ("agent") from the reference population.

Let a_{vi}, a_{vj} denote the (random) event that subject v responds positively to item i or item j, resp. Let r_v denote the number of positive answers given by subject v.

To meet the requirement that an ordering of items is "always possible", it is demanded that $a_{vi} \neq a_{vj}$, e.g. $r_v = 1$ for the two-item test (i,j); thus the order iRj is conditional on $r_v = 1$, and the probability may be written as $p(iRj \mid r_v = 1)$. Rasch's requirement of "unambiguity" is then formulated as

$$p(iRj|r_v = 1) \perp v, \tag{6}$$

which may be considered as an exact formulation of the second quotation from Rasch.

Below, in our presentation of the actual derivation, we will demonstrate that the Rasch model can be derived as the probabilistic analogue of the Guttman scalogram.

In the Guttman scalogram, the order of the dichotomous items i and j on a dimension of difficulty (denoted as iRj: i is harder than j) can be determined unequivocally in a deterministic sense when some subject v dominates j (denoted as: vSj) but does not dominate i (-vSi), assuming the domination relations comply with "composite transitivity" (Ducamp & Falmagne (1969)).

COMPOSITE TRANSITIVITY

The relation 'v dominates i', written as vSi, is observed when, e.g. object v solves problem i, or when subject v accepts statement i. Without loss of generality, the ordering can be reversed and the data interpreted as the complementary relation. Ducamp & Falmagne (1969) have formulated the following deterministic property, which we call composite transitivity. It is the same as Guttman's (1950) principle of scalability, and it defines a Guttman scale:

$$\text{if (vSi, not wSi \& wSj) then vSj} \tag{7}$$

It says that, if v dominates i, w does not dominate i, but w dominates j, then v must dominate j. From this property, it follows, as shown by Ducamp & Falmagne that a simple composite ordering of v, w,... and i, j,... exists.

Thus, by (7):

$$iRj =: vSj \,\&\, -vSi, \perp v \tag{8}$$

is a simple order, and if iRj for some v, then iRj for any v. Taking (8) as the definition of the ordering among two items, we can determine their order whenever some subject v dominates j but not i.

Such a differential response can only be observed with subjects who score $r_v = 1$ in the item pair (i,j). In such a case, the item order iRj is established independently of the subject sampled from the reference population (sample-independently):

$$(vSj \,\&\, -vSi \mid r_v = 1) \perp v. \tag{9}$$

Similarly, the ordering among two subjects is defined by

$$vQw =: vSi \,\&\, -wSi \tag{10}$$

and is observed whenever two subjects respond differently to a single item, regardless of which item that is.

PROBABILISTIC GENERALIZATION FOR K=2

Let K denote the number of items considered. For K=2, eq. (8) may be generalized to a probabilistic context as follows. The probability of the event iRj can be determined unequivocally in a stochastic sense when

$$p(iRj) =: p(vSj \,\&\, -vSi \mid r_v = 1) \perp v. \tag{11}$$

The ordering will be called a stochastically consistent item ordering.

The right part of (11) is identical to Fischer's equation (12.4.1) (1974, 196) where it is derived from the requirement of sufficiency. In effect, it is as much as a definition of sufficiency. The present derivation is, however, not based on sufficiency as a statistical requirement but on the requirement of a consistent stochastic ordering of the items, analogous to the ordering derived from composite transitivity in the scalogram structure.

From (11) we can derive:

$$
\begin{aligned}
p(vSj \;\&\; -vSi \mid r_v = 1) &= \frac{p(vSj \;\&\; -vSi)}{p(r_v = 1)} \\
&= \frac{p(vSj \;\&\; -vSi)}{p(vSj \;\&\; -vSi) + p(-vSj \;\&\; vSi)} \\
&= \frac{p(vSj)(1-p(vSi))}{p(vSj)(1-p(vSi)) + p(vSi)(1-p(vSj))} \\
&\perp v \qquad (12)
\end{aligned}
$$

assuming local stochastic independence. When we define p(vSi) as the ICC $f_i(\xi) = f_i$ of item i, (12) becomes

$$\frac{f_j(1-f_i)}{f_j(1-f_i) + f_i(1-f_j)} \perp v \qquad (13)$$

from which, assuming f to be continuous and differentiable, and nowhere equal to zero, it follows that f_i is given by the item characteristic curve of the Rasch model (Roskam (1983), 82): $f_i(\xi) = 1/\,[\,1+\exp(\sigma_i-\xi_v)]$ or

$$f_i(\theta_v) = \frac{\theta_v\varepsilon_i}{1 + \theta_v\varepsilon_i} \qquad (14)$$

where $\theta = \exp(\xi)$, and $\varepsilon = \exp(-\sigma)$. (Details of the derivation are given in the appendix). Thus the Rasch model can be derived from the requirement of stochastically consistent item orderings, and in this sense the model can be considered as the probabilistic counterpart of the scalogram: the (probability of the) ordering iRj in the test (i,j) is established independent of the selection of a subject. Since the Rasch model is symmetrical in i and v, (11) could be replaced by its analogue for a stochastically consistent subject ordering.

Note that (12) is in the Rasch model equal to

$$
\begin{aligned}
p(iRj) &=: p(vSj \;\&\; -vSi \mid r_v = 1) \\
&= p_{ij}(0,1 \mid r_v = 1,\; K=2) = \varepsilon_j/(\varepsilon_i + \varepsilon_j) \qquad (15)
\end{aligned}
$$

Eq. (15) implies that the probability p(iRj) conforms to the Luce choice rule (cf. Andrich, (1978c)), and

$$\frac{p(iRj)}{p(jRi)} = \frac{\varepsilon_j}{\varepsilon_i} \qquad (16)$$

is obviously independent of what other items are included in the test. This means that p(iRj) computed according to (15) is both independent of v and of the other items in the test.

The derivation presented above concentrates on the ordering of two items. One might consider the possibility of generalizing this derivation to multiple item orderings iRjRk... by applying stochastic transitivity on the item pair probabilities p(iRj), employing a reasoning that would be similar to Ducamp & Falmagne's (1969) treatment of composite measurement.

However, such a generalization would not be correct, because in this way a multiple ordering of probabilities would be obtained instead of the required probability of a multiple ordering. Therefore we propose in the next section a generalization of p(iRj) to multiple item orderings which is analogous to Roskam's (1983) derivation for the ordering of a pair of items.

2.1. MULTIPLE ITEM ORDERINGS

The case K=3

We first discuss the case K=3. In the Guttman scalogram, the multiple order iRjRk can be determined unequivocally only when two subjects are observed with response vectors (0, 0, 1) and (0, 1, 1). Conversely, a multiple order of subjects uQvQw is defined with the two item response vectors (0, 0, 1) and (0, 1, 1). As the argument is entirely symmetrical in subjects and items, we give the following derivations in terms of the item ordering iRjRk.

By analogy, a generalization of (11) is

$$\begin{aligned}
&p(iRjRk) =: \\
&p(vSk\ \&\ {-vSj}\ \&\ {-vSi}\ \&\ wSk\ \&\ wSj\ \&\ {-wSi} \mid r_v = 1\ \&\ r_w = 2) = \\
&p(vSk\ \&\ {-vSj}\ \&\ {-vSi} \mid r_v = 1) * p(wSk\ \&\ wSj\ \&\ {-wSi} \mid r_w = 2) \perp v. \qquad (17)
\end{aligned}$$

In the Rasch model probability (17) is indeed independent of the subjects v, w:

$$\begin{aligned}
&p(vSk\ \&\ {-vSj}\ \&\ {-vSi} \mid r_v = 1) * (p(wSk\ \&\ wSj\ \&\ {-wSi} \mid r_w = 2) =: \\
&p_{ijk}(0,0,1 \mid r_v = 1) * p_{ijk}(0,1,1 \mid r_w = 2) =
\end{aligned}$$

$$\frac{\varepsilon_k}{\varepsilon_i + \varepsilon_j + \varepsilon_k} * \frac{\varepsilon_j\varepsilon_k}{\varepsilon_i\varepsilon_k + \varepsilon_i\varepsilon_j + \varepsilon_j\varepsilon_k}\ . \qquad (18)$$

Probability (18) can be rewritten as (using (15))

$$p(iRjRk) = \frac{\varepsilon_j\varepsilon_k^2}{(\varepsilon_i+\varepsilon_j)(\varepsilon_j+\varepsilon_k)(\varepsilon_i+\varepsilon_k)+\varepsilon_i\varepsilon_j\varepsilon_k}$$

$$= \frac{\left(\frac{\varepsilon_j}{\varepsilon_i+\varepsilon_j}\right)\left(\frac{\varepsilon_k}{\varepsilon_j+\varepsilon_k}\right)\left(\frac{\varepsilon_k}{\varepsilon_i+\varepsilon_k}\right)}{1+\left(\frac{\varepsilon_j}{\varepsilon_i+\varepsilon_j}\right)\left(\frac{\varepsilon_k}{\varepsilon_j+\varepsilon_k}\right)\left(\frac{\varepsilon_i}{\varepsilon_i+\varepsilon_k}\right)}$$

$$= \frac{p(iRj)p(jRk)p(iRk)}{1+p(iRj)p(jRk)(1-p(iRk))} \qquad (19)$$

Note that p(iRj) is defined according to (15), which is independent of K. Since p(iRj) complies with the product rule

$$\frac{p(jRi)p(kRj)p(iRk)}{p(iRj)p(jRk)p(kRi)} = 1, \qquad (20)$$

we can derive from (19):

$$\frac{p(iRjRk)}{p(jRiRk)} = \frac{p(iRj)p(jRk)p(iRk)(1+p(jRi)p(kRj)p(iRk))}{p(jRi)p(iRk)p(jRk)(1+p(iRj)p(jRk)p(kRi))} = \frac{p(iRj)}{p(jRi)} \qquad (21)$$

which implies

$$p(iRj) = \frac{p(iRjRk)}{p(iRjRk)+p(jRiRk)} \qquad (22)$$

Equation (22) formulates the decomposition rule according to which the binary probability p(iRj) may be obtained from the triple probabilities p(iRjRk) in the Rasch model complying with the definitions in (11) and (17). In their overview of general probabilistic ranking theories, Luce and Suppes ((1965), 351-358) present a few examples of decomposition rules. E.g. an obvious choice seems to be

$$p(iRj) = p(iRjRk) + p(iRkRj) + p(kRiRj) \qquad (23)$$

which does not comply with (22) however. Obviously (11) and (17) as well as their generalization to rankings of K > 3, act as a ranking postulate.

Generalization of (11) and (17) to K > 3

Formula (15) can be written as*

$$p(iRj) = \frac{\gamma_1^{(i)}}{\gamma_1} \qquad (24)$$

---

* In the following equations, the γ functions are defined as follows: $\gamma_r^{(i,j)}$ is the sum of all products $\varepsilon_k\varepsilon_l\varepsilon_h\ldots$, each consisting of r terms

and (19) as

$$p(iRjRk) = \frac{\gamma_1^{(i,j)}\,\gamma_2^{(i)}}{\gamma_1\gamma_2}\,. \tag{25}$$

Note that (24) and (25) are not directly comparable since γ_1 is computed as $\varepsilon_i + \varepsilon_j$ in (24) but as $\varepsilon_i + \varepsilon_j + \varepsilon_k$ in (25).

Generalizing (17) and (18), the probability of an ordering of K items is:

$$\begin{aligned}
p(1R2R\ldots RK) =: {} & p_{12\ldots K}(0,0,\ldots,0,1 \mid r=1) \\
& * \; p_{12\ldots K}(0,0,\ldots,0,1,1 \mid r=2) \\
& * \; \ldots\ldots\ldots\ldots \\
& * \; \ldots\ldots\ldots\ldots \\
& * \; p_{12\ldots K}(0,1,1,\ldots,1 \mid r=k) \\
= {} & \frac{\gamma_1^{(1,\ldots,K-1)}}{\gamma_1} * \frac{\gamma_2^{(1,\ldots,K-2)}}{\gamma_2} * \ldots * \frac{\gamma_{K-1}^{(1)}}{\gamma_{K-1}} \\
= {} & \prod_{r-1}^{K-1} \frac{\gamma_r^{(1,\ldots,K-r)}}{\gamma_r}\,,
\end{aligned} \tag{26}$$

which is again independent of the subject sample; all γ-functions in (26) are computed on the K item parameters $\varepsilon_1,\ldots,\varepsilon_K$.

3. SOME ALTERNATIVES

In the preceding section, our main argument concentrated on a definition of the order relation among items (or subjects, resp.). It can be argued that other definitions of the order relation iRj are possible, and might lead to other models for measuring subjects and items. We discuss two such possibilities.

The first is:

$$iRj \Leftrightarrow p(vSj \;\&\; -vSi \mid r_v=1) > .50 \text{ for all } v \tag{27}$$

The second possibility is

$$iRj \Leftrightarrow p(vSj) > p(vSi) \text{ for all } v. \tag{28}$$

The latter definition is equivalent to non-intersecting item response curves.

taking all $\binom{K-2}{r}$ combinations from $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4,\ldots,\varepsilon_K$, excluding $\varepsilon_i,\varepsilon_j$. $\gamma_0=1$, by definition. The following relation holds: $\gamma_r^{(j,\ldots)} = \varepsilon_i\gamma_{r-1}^{(i,j\ldots)} + \gamma_r^{(i,j,\ldots)}$.

Both definitions do not appear to be feasible as a definition. The problem with these definitions is, that they define iRj through unknown probabilities. These definitions make the event or assessment of iRj dependent on unobservable conditions. In contrast to these definitions, our approach in the preceding section was based on an empirical definition of iRj, followed by the requirement that the probability of this event is independent of the observed subjects. Other definitions, such as in (28) have no empirical content. One can, of course, estimate the probabilities in (28), and infer iRj, but that would be a post hoc definition, and would be as good or bad as, e.g. saying that iRj if the frequency of vSj is larger than the frequency of vSi in some representative sample (with no criterium as to the representativeness of the sample).

In the context of probabilistic data, any statement concerning subjects and/or items can only be a probabilistic statement. In order to be meaningful and consistent, such probabilistic statements have to satisfy certain conditions. This was the approach adopted in the preceding section. Any derivation which does not treat iRj as a probabilistic event, would not be valid.

By this reasoning we do not, of course, exclude that models can be formulated which imply a simple ordering of subjects and items. Any holomorphic model implies such ordering. In fact, even a non-homomorphic model is compatible with a definition of an ordering of the items: e.g. in the Birnbaum model, we can define the ordering of the item by the so-called difficulty parameter, but we might as well define it by the discrimination parameter. The issue is not whether we can define an item ordering based on some model, but whether we can base a model on a proper definition of the item-ordering, taking into account that we deal with probabilistic data, and that we want probabilistic statements about items independent of the subject sample.

4. SPECIFIC OBJECTIVITY AS A METHODOLOGICAL REQUIREMENT

The derivation in section 2 shows that a specifically objective model produces a consistent subject order with respect to some construct or trait. Subject measurement is, in a stochastic sense, independent of the specific set of items selected from the universe of content defining the pertinent behavioral domain. A consequence of this is that a specifically objective measurement model is the only meaningful instrument for defining and measuring a theoretical construct supposed to underly behavior in that domain.

Before we elaborate this 'meaningfulness argument', we have to refute one possible counter-argument.

It may be that in the behavioral sciences, specific objective measurement can be established far less frequently than in physical sciences. In general, the Rasch model imposes rather strong requirements on the data. This is, however, an empirical argument, and as such cannot match the methodological status of the principle of specific objectivity. A methodology is a program in the first place, a prescription for obtaining meaningful results from empirical research. Empirical data either are obtained under this program - and they do not refute it - or they are not obtained under it, in which case inferences drawn from them can simply not be considered scientifically meaningful. Specific objectivity can only be refuted by other methodological principles which, essentially, argue why it would not make sense. An example of such an argument might be that a principle that never works because it is not adequate in view of the nature of the phenomena under study, should be discarded. At present, it seems too early for such a conclusion.

Statements about subjects are general only when they are independent of the specific measurement 'probe', 'agent', or 'item' sampled from the pertinent domain or universe of content. If a general trait underlies the responses of subject to items or probes from that universe of content, then any sample of items from that universe of content should be equally valid to measure the theoretical construct that is addressed by them. If, for instance, an attitude like 'conservatism' is to be meaningful at all, and is defined with respect to a defined domain of social issues, we should be able to measure subjects' conservatism by observing their responses to items from that domain, and the comparative inference concerning two people's degree of conservatism should not depend on the items sampled from that domain. For, if it were dependent on the sample of items, one could as well conclude that one person is more conservative than the other, as the other way around. If that were the case - which of course is a matter to decide empirically - the construct of 'conservatism' would not be a unitary, unidimensional construct, but a combination or amalgation of several traits or dispositions, making the very concept of 'conservatism' evanescent, and not meaningful as a trait with respect to which people can be ordered. The ordering would depend on the items sampled to assess it. Of course, specific objectivity and sample independence should not be confused with the reliability and the discriminative power of a certain sample of items administered to a particular group of subjects. E.g. items which, though belonging to a homogeneous universe of conservatism items and satisfying the Rasch model, are rejected by (almost) everybody, would not be much good in a conservatism questionnaire, except with the most conservative subjects.

In summary, we find that specific objectivity, and hence the Rasch model, is a basic methodological condition for the definition and measurement of theoretical constructs (e.g. verbal ability, internal control, machiavelism). It is of special importance for the theoretical study of attitudes and personality traits, where it could be very helpful isolating and measuring such constructs. To what extent the requirement of specific objectivity can actually be met, is a matter for empirical research to decide. In that way, it is also a test for the meaningfulness of personality constructs.

APPENDIX

Assuming that f_i, f_j, ... are continuously differentiable and nowhere equal to zero, the reciprocal of eq. (13) can be written as:

$$\frac{d}{d\xi}\,\frac{f_i(1-f_j)}{f_j(1-f_i)} = 0 \qquad (29)$$

Expanding and simplifying leads to:

$$f_j'(f_i-f_i^2) - f_i'(f_j-f_j^2) = 0;$$

$$\frac{f_j'}{f_j(1-f_j)} = \frac{f_i'}{f_i(1-f_i)}$$

So, for every i,j, we have:

$$\frac{f_i'}{f_i(1-f_i)} = \text{a constant, independent of } \xi.$$

This differential equation is solved by

$$f_i(\xi) = \frac{1}{1+\exp[-c(\xi-\sigma_i)]} \qquad (30)$$

where c is an arbitrary constant.

REFERENCES

[1] Andrich, D., Relationships between Thurstone and Rasch approaches to item scaling, Applied Psychological Measurement, 2 (1978), 449-460.

[2] Douglas, G.A., Conditional inference in a generic Rasch model, Invitational Seminar on the Improvement of Measurement in Education and Psychology, Melbourne (1980), 22-23 May.

[3] Ducamp, A. & Falmagne, J.-C., Composite measurement, Journal of Mathematical Psychology, 6 (1969), 359-390.

[4] Fischer, G.H., Einführung in die Theorie psychologischer Tests, Huber, Wien (1974).

[5] Fischer, G.H., & Pendl, P., Individualized testing on the basis of the dichotomous Rasch model, In L.J.Th. van der Kamp, W.F. Langerak & D.N.M. de Gruijter, Psychometrics for educational debates, New York, Wiley (1980), 171-188.

[6] Guttman, L., Chapters 2, 3, 6, 8 and 9 in S. Stouffer et al. (eds.), Measurement and Prediction, Princeton, University of Princeton Press (1950).

[7] Jansen, P.G.W., Rasch analysis of attitudinal data, Doctoral Dissertation, University of Nijmegen, Print: Rijks Psychologische Dienst, The Hague, The Netherlands (1983).

[8] Krantz, D.H., Luce, R.D., Suppes, R., & Tversky, A., Foundations of measurement, Vol. 1, Additive and polynomial representation, New York, Academic Press (1971).

[9] Luce, R.D., & Suppes, P., Preference, utility and subjective probability, In R.D. Luce, R.R. Bush, & E. Galanter, Handbook of Mathematical Psychology, Vol. III, New York, Wiley (1965).

[10] Ramsay, J.O., Review of foundations of measurement by David H. Krantz et al., Psychometrika, 40 (1975), 257-262.

[11] Rasch, G., Probabilistic models for some intelligence and attainment tests, Copenhaguen: Danish Institute for Educational Research (1960), reprinted (1980).

[12] Rasch, G., An informal report on a theory of objectivity in comparisons, In Proceedings of the NUFFIC International summer session in science at 'Het Oude Hof', The Hague, 14-28 July,(1966).

[13] Rasch, G., On specific objectivity: an attempt at formalizing the request for generality and validity of scientific statements, Danish Yearbook of Philosophy, 14 (1977), 58-94.

[14] Roskam, Edw. E., Methodenleer, In H.C.J. Duijker & P.A. Vroon (red.) Codex Psychologicus, Amsterdam, Elsevier (1981), 31-68.

[15] Roskam, Edw. E., Allgemeine Datentheorie, In H. Feger & H. Bredenkamp (Ed.), Messen und Testen; Band 3 der Serie Forschungsmethoden der Psychologie der Enzyklopädie der Psychologie, Göttingen, Hogrefe (1983) 1-135.

[16] Scheiblechner, H., Specifically objective stochastic latency mechanism, Journal of Mathematical Psychology, 19 (1979), 18-38.

[17] Spada, H., & Kempf, W.F., (Eds.), Structural models of thinking and learning, Bern, Huber (1977).

The paper is based on Roskam (1983), and on a chapter from the doctoral dissertation of the first author (Jansen (1983)).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

A DEFINITION OF PARTIAL INTERVAL ORDERS

Marc Roubens
Faculté Polytechnique de Mons, Belgium
Philippe Vincke
Université Libre de Bruxelles, Belgium

Preference modelling often uses concepts such as preorders, semiorders and interval orders and, in most cases, it is assumed that the relations are complete (no incomparability).
However, incomparabilities naturally appear in many decision problems, as for example, in multicriteria problems. That is the reason why the study of partial relations becomes more and more important.
In this paper, we propose a definition of partial interval order which extends the concept of interval order, which is compatible with a numerical representation and which is defined by properties very easy to verify.

1. INTRODUCTION

Preference modelling uses the concepts of preorder, semiorder, interval order, pseudo-order and fuzzy order to a large extent. In most cases, it is assumed that the relations are complete (no incomparability).

When one is confronted with outranking relations illustrating in the best way possible the preference or the global judgment on the set A of potential actions, one is in fact provided with incomplete relations. Defining partial interval orders is consequently worth doing in the relatively current cases when one meets only with relations of preference (P), indifference (I) and incomparability (R).

In this paper we propose a definition of partial interval order which extends the concept of interval order, which is compatible with a numerical representation and which is defined by properties very easy to verify.

2. NUMERICAL REPRESENTATIONS OF COMPLETE AND PARTIAL PREORDERS

It is well known that, given an asymmetric relation P and a reflexive and symmetric relation I on a finite set A, the couple (P,I) is a complete preorder iff

a) $\forall\ a,b \in A$: $a\ P\ b$ or $b\ P\ a$ or $a\ I\ b$ (exclusive or)
b) P is transitive
c) I is transitive:

iff a function g exists such that, $\forall\ a,b \in A$:

$$\left\{\begin{array}{l} a\ P\ b \Leftrightarrow g(a) > g(b), \\ a\ I\ b \Leftrightarrow g(a) = g(b). \end{array}\right.$$

On the other hand, (P,I) is a partial preorder iff the conditions b) and c) are satisfied and if (P,I) is a partial preorder, then a function g exists such that, $\forall\ a,b \in A$:

$$\left\{\begin{array}{l} a\ P\ b \Rightarrow g(a) > g(b), \\ a\ I\ b \Leftrightarrow g(a) = g(b). \end{array}\right.$$

So, the concept of partial preorder is sufficient but not necessary to have the last numerical representation, because this representation does not imply the transitivity of P. In fact, the necessary and sufficient condition to have this numerical representation is that the relation P does not contain any circuit and that I is transitive. (P contains a circuit if there exist elements $a_1, a_2, \ldots, a_n$ such that $a_i\ P\ a_{i+1}$, $\forall\ i$ and $a_n\ P\ a_1$). This is an immediate consequence of a well-known theorem of graph theory (see [4]).

3. SOME CHARACTERIZATIONS OF THE INTERVAL ORDERS

Let again P be an asymmetric relation and I a reflexive and symmetric relation defined on a finite set A. Let R, S_+ and S_- be the relations defined as follows:

$$\left\{\begin{array}{l} a\ R\ b \text{ iff } a \not P b,\ b \not P a \text{ and } a \not I b, \\ a\ S_+\ b \text{ iff } \exists\ c\text{: } a\ P\ c \text{ and } c\ I\ b, \\ a\ S_-\ b \text{ iff } \exists\ d\text{: } a\ I\ d \text{ and } d\ P\ b. \end{array}\right.$$

Let $G = (A,U)$ be the graph associated to (P,I) as follows:

$$\left\{\begin{array}{l} a\ P\ b \Leftrightarrow (a,b) \in U \text{ and } (b,a) \notin U \text{ (P-type arc)}, \\ a\ I\ b \Leftrightarrow (a,b) \in U \text{ and } (b,a) \in U \text{ (I-type arc)}. \end{array}\right.$$

Let us call a circuit of G a set of couples $\{(a_i, a_{i+1}), i=1,\dots,k\}$ of elements of A such that

$$\begin{cases} a_{k+1} = a_1, \\ a_i \; P \; a_{i+1} \text{ or } a_i \; I \; a_{i+1}, \; \forall \, i, \end{cases}$$

the length of the circuit being k.

THEOREM 1

If $R = \emptyset$, the following properties are equivalent.

(i) $\forall$ a,b,c,d $\in$ A : a P b, b I c, c P d $\Rightarrow$ a P d (which is denoted $P \, I \, P \subset P$);

(ii) In G, every circuit of length 4 contains at least two consecutive I-type arcs;

(iii) In G, every circuit contains at least two consecutive I-type arcs;

(iv) The relation S_- does not contain any circuit;

(v) The relation S_+ does not contain any circuit;

(vi) There exist two functions g and q such that, $\forall$ a,b $\in$ A:

$$a \; P \; b \Leftrightarrow g(a) > g(b) + q(b),$$

$$a \; I \; b \Leftrightarrow \begin{cases} g(a) \leq g(b) + q(b), \\ g(b) \leq g(a) + q(a), \end{cases}$$

$$q(a) \geq 0.$$

These properties define what is usually called an interval order ([1], [2]).

Proof:

(i) $\Rightarrow$ (ii):

Let (a,b), (b,c), (c,d), (d,a) $\in$ U; if this circuit does not contain two consecutive I-type arcs, then we have a P b and c P d (or every similar situation with other couples) which implies a P d (consequence of (i) and of the fact that b P c or b I c); this is in contradiction with (d,a) $\in$ U.

(ii) $\Rightarrow$ (iii):

If a circuit of G does not contain two consecutive I-type arcs, then it contains elements a,b,c,d such that a P b, b I c, c P d, which implies, by (ii), a P d; this defines another circuit of G, the length of which is smaller than the length of the initial circuit.

Repeating this operation again and again, we obtain a contradiction with (ii).

(iii) $\Rightarrow$ (iv): evident.

(iv) $\Rightarrow$ (v): evident.

(v) $\Rightarrow$ (vi):

Since S_+ has no circuit, a function g exists such that, $\forall\, a,b \in A$:

$$a\ S_+\ b \Rightarrow g(a) > g(b) \qquad (\text{cf. §2}).$$

To each element a of A, we associate the number q(a) such that

$$\begin{cases} g(a) + q(a) < g(b),\ \forall\, b\colon b\ P\ a, \\ g(a) + q(a) \geq g(c),\ \forall\, c\colon c\ I\ a; \end{cases}$$

this is always possible because b P a and c I a imply $b\ S_+\ c$, thus $g(b) > g(c)$; moreover $q(a) \geq 0$ because a I a.

We obtain so two functions g and q such that, $\forall\, a,b \in A$:

$$\begin{cases} a\ P\ b \Rightarrow g(a) > g(b) + q(b), \\ a\ I\ b \Rightarrow \begin{cases} g(a) \leq g(b) + q(b), \\ g(b) \leq g(a) + q(a); \end{cases} \end{cases}$$

as $R = \emptyset$, the simple implications may be replaced by double implications.

(vi) $\Rightarrow$ (i):

a P b, b I c and c P d imply

$g(a) > g(b) + q(b) \geq g(c) > g(d) + q(d)$, so that a P d. Q.E.D.

4. TOWARDS A DEFINITION OF PARTIAL INTERVAL ORDER

In terms of preferences, an interval order assumes that given two elements of A, you must have preference or indifference between them. We want now to define a concept which allows incomparabilities ($R \neq \emptyset$) and which extends the concept of interval order in the same way as a partial preorder extends a complete preorder. In this purpose, it is interesting to see which are the connections between the six properties of theorem 1 when $R \neq \emptyset$. Note that in this case, property (vi) must be replaced by

(vi)' There exist two functions g and q such that, $\forall\, a,b \in A$:

$$\begin{cases} a\ P\ b \Rightarrow g(a) > g(b) + q(b), \\ a\ I\ b \Rightarrow \begin{cases} g(a) \leq g(b) + q(b), \\ g(b) \leq g(a) + q(a), \end{cases} \\ q(a) \geq 0. \end{cases}$$

THEOREM 2

If $R \neq \emptyset$, then

1°) (i) ⇒ (iii) ⇔ (iv) ⇔ (v) ⇔ (vi)' ⇒ (ii)

2°) (iii) $\not\Rightarrow$ (i)

3°) (ii) $\not\Rightarrow$ (iii).

Proof:

1°) It is evident that (i) ⇒ (iv) and that (iii) ⇒ (iv) ⇒ (v).
The proof of (v) ⇒ (vi)' is similar to the proof of (v) ⇒ (vi) but stops before the use of the fact that $R = \emptyset$.
The fact that (vi)' ⇒ (iii) is very easy to verify, as well as the fact that (vi)' ⇒ (ii).

2°) Consider a P b, b I c, c P d, a R d, a R c and b R d: this example shows that (iii) $\not\Rightarrow$ (i).

3°) Consider a P b, b I c, c P d, d I e, e P f, f I a, a R c, a R d, a R e, b R d, b R e, b R f, c P e, c R f, d R f: this example shows that (ii) $\not\Rightarrow$ (iii).
Q.E.D.

We see that several ways exist for the definition of a partial interval order. Nevertheless, given the importance of the numerical representation of a preference structure, we think that property (iii) (or (iv) or (v)) must be imposed. On the other hand, it is known that the passage of the complete preorder to the interval order is performed to the detriment of the transitivity of I while the transitivity of P is maintained. This is the reason why we propose to maintain the transitivity of P in the definition of a partial interval order.

This does not mean that a structure where P is not transitive does not present any interest, but we think that such a structure must have another name. These considerations lead to the following definition.

DEFINITION

(P,I) is a partial interval order iff

1°) the relation S_+ (or S_-) does not contain any circuit;

2°) P is transitive.

REMARK

The choice of (v) instead of (iii) (which is equivalent) in the definition is due to algorithmic considerations: it is much easier to build S_+ and to verify the absence of circuit in it (algorithms exist in graph theory) than to verify property (iii).

5. CONCLUSIONS

We have proposed in this paper a definition of partial interval order, which has the following characteristics:

- if $R = \emptyset$ (no incomparability), the partial interval order is an interval order;
- the conditions defining a partial interval order are very easy to verify;
- the numerical representation of a partial interval order is similar to the representation of an interval order: the difference consists in the replacement of double implications by simple implications, as in the passage from complete to partial preorders;
- every partial interval order (P,I) can be completed in an interval order (P',I') such that $P \subset P'$ and $I \subset I'$, by replacing in a judicious way incomparabilities by preferences or indifferences; in fact, we have only to put

$$\begin{cases} a\,P'\,b \Leftrightarrow g(a) > g(b) + q(b), \\ a\,I'\,b \Leftrightarrow \begin{cases} g(a) \leq g(b) + q(b), \\ g(b) \leq g(a) + q(a), \end{cases} \end{cases}$$

where g and q are the functions defined in (vi)'.

REFERENCES

[1] Fishburn, P.C., Intransitive indifference with unequal indifference intervals, Journal of Mathematical Psychology, 7 (1970), 144-149.

[2] Roberts, F.S., On non transitive indifference, Journal of Mathematical Psychology, 7 (1970), 243-258.

[3] Roubens, M., Analyse et agrégation des préférences: modélisation, ajustement et résumé de données relationnelles, Revue Belge de Stat., d'Informatique et de Recherche Opérationnelle, 20 (1980), 36-67.

[4] Roy, B., Algèbre moderne et théorie des graphes, 2 vol., Dunod, (1969).

[5] Roy, B., Préférence, indifférence, incomparabilité: relations binaires et structures fondamentales, document 9, LAMSADE, Université Dauphine, Paris (1980).

[6] Vincke, Ph., Vrais, quasi, pseudo et précritères dans un ensemble fini: propriétés et algorithmes, Cahiers du LAMSADE, Université Paris IX Dauphine (1980).

[7] Vincke, Ph., Preference modelling: a survey and an experiment, Operational Research 81, J.P. Brans (ed.), North-Holland (1981).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

CAUSAL LINEAR STOCHASTIC DEPENDENCIES: THE FORMAL THEORY

Rolf Steyer
University of Trier
Trier, Federal Republic of Germany

The formal background of the theory of causal linear stochastic dependence is provided, which was introduced by Steyer (1984). The theory presented is concerned with those kinds of dependencies which can be described by specifying the functional form of a conditional expectation $E(Y|X)$. This includes also those situations in which X is a multidimensional random variable. The main concepts of the theory are *causal* and *weak causal linear stochastic dependencies*, the definition of which is based on the pre- and equiorderedness relations of sigma-fields and stochastic variables, on the notion of *potential disturbing sigma-fields and variables*, as well as on the invariance and on the average conditions. These concepts are formally defined and their properties are studied in some detail. *Causal* linear stochastic dependence is defined by the preorderedness condition that the influencing variable is antecedent to the influenced variable and by the *invariance condition*, whereas *weak causal* linear stochastic dependence is defined by the preorderedness and *average conditions*. Both, the invariance and the average conditions, and therefore both kinds of causal hypotheses, can empirically be tested in experimental as well as in nonexperimental observational studies.

1. INTRODUCTION

"Correlation does not prove causality", is a statement univocally found in textbooks on applied statistics. However, whenever one looks for a definition of causal dependence, one either finds treatments on philosophical theories only loosely related to the concepts of correlation, regression, and the experimental control techniques such as randomization (see e.g. Bagozzi (1980) or Heise (1975)), or treatments of experimental and quasi-experimental control techniques (see e.g. Cook & Campbell (1979)). Although these discussions are very useful and instructive in many respects, there is no formal theoretical connection between the correlation, regression, and analysis of variance models applied and the control techniques discussed. What is, in formal terms, the difference between an analysis of variance model in a purely observational study at one hand and in a randomized experiment at the other?

Steyer (1983, 1984) has taken some first steps to bridge the gap between ideas of causal dependence, stochastic models, and experimental control techniques. The basic question raised is, which are the formal properties that make a stochastic model a causal one? Restricting the discussion to linear stochastic dependencies, i.e. to those dependencies which can be described by conditional expectations, Steyer (1984) proposed such properties defining two types of nonspurious or causal linear stochastic dependencies, a weak and a strong one, each of which are distinguished from non-causal linear stochastic dependencies by two conditions. Beside the *preorderedness condition* that the influencing variable X is antecedent to the influenced variable Y, the crucial condition for *causal linear stochastic dependence* of a random variable Y on a, possibly multidimensional, random variable X is the *invariance condition* postulating that $E(Y|X,W)$ as= $E(Y|X) + H \circ W$ holds for all potential disturbing variables W, where $H \circ W$ is a W-measurable composition of a function H with W. The invariance condition states that, if any potential disturbing variable W is included as another conditioning variable , it only adds to $E(Y|X)$ in the equation for $E(Y|X,W)$, leaving $E(Y|X)$ invariant.

The crucial condition for *weak causal linear stochastic dependence*, on the other hand, is the *average condition* postulating that, for all potential disturbing variables W, $E(Y|X=x) = \int E(Y|X=x, W=w)\, P_W(dw)$ holds, for P_X-almost all $E(Y|X=x)$. This condition means that $E(Y|X=x)$ is the average of the conditional expectations $E(Y|X=x, W=w)$ across the values w of W, for P_X-almost all $E(Y|X=x)$. If W is a discrete stochastic variable, this equation can be written:

$$E(Y|X=x) = \Sigma_W E(Y|X=x, W=w) \cdot P(W=w).$$

It will be shown that this average condition holds, for example, if X and all potential disturbing variables W are stochastically independent, a condition which may be assumed to hold in randomized experiments, where X indicates group membership and the potential disturbing variables W represent properties of the experimental units (subjects) before or at the time of the treatment. This provides the desired link to the experimental control techniques such as randomization and matching.

In the formulation of the preorderedness, invariance and average conditions above, formally undefined terms have been used that have to be eliminated in order to construct a formal theory consisting only of mathematically well-defined terms. So far, we have no formal criteria to decide whether or not a specified stochastic variable X is antecedent to a specified stochastic variable Y, and whether or not a specified stochastic variable W is a potential disturbing variable. Can $W_1 = X^2$ be a potential disturbing variable with respect to the dependence of Y on X? Can $W_2 = X+Z$ be a potential disturbing variable? Are variables mediating between X and Y potential disturbing variables, or variables which are influenced by Y?

In the following sections, we first summarize some basic concepts of probability theory and introduce notational conventions. Then, we give a formal definition of the pre- and equiorderedness relations and study their properties. Next, we define potential disturbing sigma-fields and variables, which will make the average and invariance conditions discussed above well defined formal concepts, too. Then, we give formal definitions of causal and weak causal linear stochastic dependencies, and investigate the properties of these concepts.

2. SOME BASIC CONCEPTS OF PROBABILITY THEORY AND NOTATION

In this section, a brief summary and notational conventions of some basic concepts of probability theory are given, which seem to be essential for a proper understanding of the theory proposed. For detailed introductions, the reader is referred to Bauer (1974), Breiman (1968), Gänsler and Stute (1977), Halmos (1969), or Loève (1977, 1978).

The fundamental assumption of every stochastic substantive (i.e. psychological, sociological, etc.) model is that the experiment, or more generally, the part of reality to be described, can be represented by a probability space, the definition of which is based on the following concepts.

Let $\mathbb{N} = \{1,2,\dots\}$ be the set of natural numbers. A *sigma-field* $\mathcal{A}$ on Ω is defined to be a set of subsets of Ω with the following three properties. (a) $\Omega \in \mathcal{A}$. (b) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$, where $A^c := \Omega - A$ denotes the complement of A. (c) If $(A_i,\ i \in \mathbb{N})$ is a sequence of elements A_i of $\mathcal{A}$, then their union $\cup_{i \in \mathbb{N}} A_i$ is an element of $\mathcal{A}$.

The intersection of a family $(\mathcal{A}_i,\ i \in I)$ of sigma-fields $\mathcal{A}_i$ on Ω is also a sigma-field on Ω. If $\mathcal{E}$ is a set of nonempty subsets of Ω, then *the sigma-field $\mathcal{A}(\mathcal{E})$ generated by the set system $\mathcal{E}$* is defined to be the intersection of all those sigma-fields on Ω, which contain $\mathcal{E}$ as a subset.

A *measurable space* is defined to be a pair $(\Omega,\mathcal{A})$ of a set Ω and a sigma-field $\mathcal{A}$ on Ω.

Let $(\Omega,\mathcal{A})$ be a measurable space. A *probability measure* $P\colon \mathcal{A} \to [0,1]$ is defined to be a function assigning each $A \in \mathcal{A}$ a real nonnegative number with the following three properties. (a) $P(\emptyset) = 0$. (b) If $(A_i,\ i \in \mathbb{N})$ is a sequence of elements of $\mathcal{A}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$, then $P(\cup_{i \in \mathbb{N}} A_i) = \Sigma_{i \in \mathbb{N}} P(A_i)$. (c) $P(\Omega) = 1$.

A *probability space* can now be defined to be a triple $(\Omega,\mathcal{A},P)$ of a set Ω called the set of *elementary events*, a *sigma-field* $\mathcal{A}$ of subsets of Ω, where the elements $A \in \mathcal{A}$ are called *events*, and a *probability measure* $P\colon \mathcal{A} \to [0,1]$ which assigns the *probability* $P(A)$ to each event $A \in \mathcal{A}$.

Let $(\Omega,\mathcal{A})$ and $(\Omega',\mathcal{A}')$ be measurable spaces. A mapping $X\colon \Omega \to \Omega'$ is called *$(\mathcal{A},\mathcal{A}')$-measurable*, iff, for all $A' \in \mathcal{A}'$,

$$X^{-1}(A') := \{\omega \in \Omega\colon X(\omega) \in A'\} \in \mathcal{A}.$$

The set $X^{-1}(\mathcal{A}') := \{X^{-1}(A')\colon A' \in \mathcal{A}'\}$ is called the *sigma-field generated by X and $\mathcal{A}'$*. $X^{-1}(\mathcal{A}')$ may also be denoted by $\mathcal{A}(X,\mathcal{A}')$. If X is real-valued and $N(I)$-dimensional, $\mathcal{A}(X,\mathcal{A}') = X^{-1}(\mathcal{A}')$ is also denoted by $\mathcal{A}(X)$ and $\mathcal{A}'$ is understood to be the *Borel sigma-field* $\mathcal{B}_{N(I)}$ on $\mathbb{R}^{N(I)}$, which is defined to be the sigma-field generated by the set system of all open intervalls of $\mathbb{R}^{N(I)}$, $N(I) \in \mathbb{N}$.

Let $(\Omega_i, \mathcal{A}_i)$, $i \in I$, be a finite or infinite sequence of measurable spaces. A sigma-field generated by a sequence $(X_i,\ i \in I)$ of $(\mathcal{A}, \mathcal{A}_i)$-measurable mappings $X_i\colon \Omega \to \Omega_i$ is defined to be the sigma-field generated by the union of the sigma-fields $X_i^{-1}(\mathcal{A}_i)$, i.e.

$$\mathcal{A}[(X_i, \mathcal{A}_i),\ i \in I] := \mathcal{A}[\cup_{i \in I} X_i^{-1}(\mathcal{A}_i)] .$$

If $I = \{1, \ldots, N(I)\}$ is finite and $\Omega_i = \mathbb{R}$, for all $i \in I$, we also use the alternative notation $\mathcal{A}(X_1, \ldots, X_{N(I)})$ instead of $\mathcal{A}[(X_i, \mathcal{A}_i),\ i \in I]$.

Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces. An $(\Omega', \mathcal{A}')$-stochastic variable X on the probability space $(\Omega, \mathcal{A}, P)$ is defined to be a mapping $X\colon \Omega \to \Omega'$ that is $(\mathcal{A}, \mathcal{A}')$-measurable. A $N(I)$-dimensional real-valued stochastic variable, for example, is a $(\mathbb{R}^{N(I)}, \mathcal{B}_{N(I)})$-stochastic variable.

The measurability condition in the definition of a $(\Omega', \mathcal{A}')$-stochastic variable implies that

$$X^{-1}(A') := \{\omega \in \Omega\colon X(\omega) \in A'\} \in \mathcal{A}, \text{ for all } A' \in \mathcal{A}' .$$

This allows to define the distribution $P_X\colon \mathcal{A}' \to [0,1]$ of a $(\Omega', \mathcal{A}')$-stochastic variable X by

$$P_X(A') = P[X^{-1}(A')] , \text{ for all } A' \in \mathcal{A}' .$$

The distribution P_X of X is a probability measure on $\mathcal{A}'$.

3. PRE- AND EQUIORDEREDNESS

A necessary condition for a stochastic dependence of Y on X to be causal is that X is antecedent or, synonymously, preordered to Y. In experiments, for example, the treatment variables are manipulated before the effects on the dependent variables are assessed, and in nonexperimental studies, too, the influencing variables have to be antecedent to the influenced variables, if a causal statement should make sense at all. Even reciprocal causal dependence can much better be thought of as a process of mutual influencing of the variables involved (see e.g. Steyer (1982)), where, at each point of the process, the causing variable is preordered to the caused one. Thus, representing reciprocal causal relations by dynamic models not only allows the preservation of asymmetry of influence (see e.g. Simon (1952)) for reciprocal causality, but is also much more congruent with the dynamic nature of reality. Preorderedness of a stochastic variable X to a second one, Y, is the first concept to be defined which draws on the fact that (Ω, A, P) is assumed to represent a process.

A (Ω', A')-stochastic process on a probability space (Ω, A, P) is defined to be a family $(X_t, t \in T)$ of (Ω', A')-stochastic variables on (Ω, A, P). Hence, one might think of defining X_s to be preordered to X_t, if $s < t$, $s, t \in T$. However, such a concept would be too restricted. Oftentimes, variables have to be considered that are defined, for example, by $X := X_s \cdot X_t$, $s \neq t$. Obviously, the definition of preorderedness considered above could not be applied to X, because not only one but two points of time are involved in the definition of X.

These problems are not only of theoretical, but also of much practical interest. If, for example, the effect of a psychological treatment (therapy, training, etc.) is to be investigated, such a treatment usually extends over several sessions. Hence, if we let T be the index set of the sessions, then the treatment variable, indicating whether or not a person receives a treatment, cannot be assigned to one single $t \in T$. The same problem occurs, if we let T denote the continuous time covered by one session. An intervention of the psychologist cannot be assigned to a single $t \in T$ in the way outlined above. Similarly, repeated events, such as being cheated repeatedly by one's mother, may be a cause of someone being distrustful. Again, the causing variable cannot be assigned to a single point of time, a problem which Suppes (1970) seems to neglect (cf. Stegmüller (1983) p. 602)).

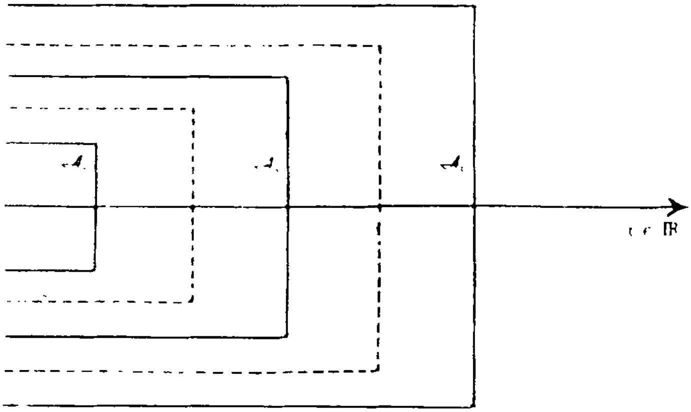


Figure 1. A monotonically increasing family $(A_t, t \in T)$ of sigma-fields.

In order not to run into these difficulties, we basically follow the approach proposed by Steyer (1983), which is, however, corrected in some points and studied in more detail. This approach is not based on a process $(X_t, t \in T)$, but on a monotonically increasing family $(A_t, t \in T)$ of sigma-fields, (see Figure 1). Let T be a subset of the set $\mathbb{R}$ of real numbers. A monotonically increasing family $(A_t, t \in T)$ of sigma-fields is defined to be a family of sigma-fields A_t on a set Ω, $t \in T \subset \mathbb{R}$, with the property that, if $s \leq t$, $s,t \in T$, then $A_s \subset A_t$. Such a monotonically increasing family of sigma-fields is obtained, for example, if we define A_t to be the sigma-field generated by all X_s, $s \in T$, $s \leq t$, where $(X_t, t \in T)$ is assumed to be a real-valued stochastic process. However, we do not need a stochastic process $(X_t, t \in T)$ to construct a monotonically increasing family $(A_t, t \in T)$ of sigma-fields, as is shown in the following example.

3.1. EXAMPLE

Consider an experiment in which two coins, each having one metal and one plastic side, are tossed onto a plate having the properties of an electro magnet which is on or off while the two coins are tossed. (For a more detailed description of this example, see Steyer, (1984)). In this application, we may choose Ω to be the set product of the sets $\Omega_1 = \{a_1,a_2\}$, $\Omega_2 = \{b_1,b_2\}$, and $\Omega_3 = \{c_1,c_2\}$. Hence, each element (a_i,b_j,c_k) of Ω denotes one of the eight elementary events that Coin 1 shows side i, Coin 2 shows side j, and the electro magnet is off (k=1) or on (k=0). Furthermore, we choose the sigma-field A on Ω to be the set of all subsets of Ω. If X denotes the stochastic variable indicating the state of the electro magnet and the variables Y_i, i=1,2, indicate the outcome of tossing coin i, a monotonically increasing family $(A_t, t \in T)$ of sigma-fields is easily obtained if we define $T := \{1,2\}$, $A_1 := A(X)$, and $A_2 := A(X,Y_1,Y_2)$, where $A(X,Y_1,Y_2)$ denotes the sigma-field generated by the real-valued stochastic variables X, Y_1, Y_2.

The example above, involving only three variables X, Y_1 and Y_2 and their generated sigma-fields, is very simple. However, it will help to illustrate the concept of preorderedness that is defined as follows.

3.2. DEFINITION

Let (Ω,A,P) be a probability space, let T be a subset of the set $\mathbb{R}$ of real numbers, let $(A_t,\ t \in T)$ be a monotonically increasing family of sigma-fields with $A(\cup_{t\in T} A_t) \subset A$, let X, W be (Ω',A')-, (Ω'',A'')-stochastic variables on (Ω,A,P), respectively, and finally, let $\mathcal{C}, \mathcal{D} \subset A$ be two sigma-fields.

(i) We say that $\mathcal{C}$ is preordered to $\mathcal{D}$ with respect to $(A_t,\ t \in T)$, iff ('iff' is an abreviation for 'if and only if')

(a) there is a $s \in T$, with $\mathcal{C} \subset A_s$, $\mathcal{D} \not\subset A_s$ and

(b) there is an element $t \in T$, $t > s$, with $\mathcal{D} \subset A_t$, $\mathcal{D} \not\subset A_s$, for all $s \in T$, $s < t$.

(ii) We say that X is preordered to W with respect to $(A_t,\ t \in T)$, iff Conditions (a) and (b) hold with

(c) $\mathcal{C} = X^{-1}(A')$, $\mathcal{D} = W^{-1}(A'')$.

If no confusion is possible, the explicit reference to $(A_t,\ t \in T)$ may be omitted. According to Conditions (a) and (b) of Definition 3.2, there is a smallest $t \in T$ for which $\mathcal{D}$ is a subset of A_t and that there is an element $s \in T$, $s < t$, such that $\mathcal{C}$ is a subset of A_s. In the example discussed above, where all variables involved are real-valued (i.e. $\Omega' = \Omega'' = \mathbb{R}$ and $A' = A'' = \mathcal{B}$, where $\mathcal{B}$ denotes the Borel sigma-field on $\mathbb{R}$), X is preordered to Y_1 and Y_2, because $A(X) := X^{-1}(\mathcal{B}) \subset A_1$, whereas $A(Y_1) := Y_1{}^{-1}(\mathcal{B}) \not\subset A_1$ and $A(Y_2) := Y_2{}^{-1}(\mathcal{B}) \not\subset A_1$, but $A(Y_1), A(Y_2) \subset A_2$. In this example, Y_1 and Y_2 are equiordered in the following sense.

3.3. DEFINITION

Let the presumptions and notations of 3.2 be valid.

(i) We say that $\mathcal{C}$ and $\mathcal{D}$ are equiordered with respect to $(A_t,\ t \in T)$, iff

(a) there is an element $t \in T$ with $\mathcal{C}, \mathcal{D} \subset A_t$ and

(b) there is no element $s \in T$, $s < t$, with $\mathcal{C} \subset A_s$ or $\mathcal{D} \subset A_s$.

(ii) We say that X and W are equiordered with respect to $(A_t,\ t \in T)$, iff Conditions (a) and (b) hold with

(c) $\mathcal{C} = X^{-1}(A')$ and $\mathcal{D} = W^{-1}(A'')$.

According to Conditions (a) and (b) of Definition 3.3, there is a smallest element $t \in T$ with $\mathcal{C}$ and $\mathcal{D}$ being both subsets of A_t, and there is no element $s \in T$, $s < t$, such that $\mathcal{C}$ or $\mathcal{D}$ are subsets of A_s.

It is easily seen that the diagram of any recursive path analysis model can be translated into a monotonically increasing family of sigma-fields so that the concepts of pre- and equiorderedness can be applied. For the path diagram of Figure 2, for example, we may define $T := \{1,2,3\}$, $A_1 := A(Z_1,Z_2)$,

$A_2 := A(Z_1,Z_2,Z_3)$, and $A_3 := A(Z_1,\dots,Z_4)$. Obviously, Z_1 and Z_2 are equiordered, Z_1 and Z_2 are both preordered to Z_3 and Z_4, and Z_3 is preordered to Z_4 with respect to $(A_t,\ t \in T)$ (see Definitions 3.2 and 3.3).

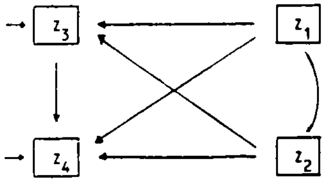


Figure 2. Path diagram for a recursive model with four variables.

We now treat some formal properties of the pre- and equiorderedness relations. The first one is that preorderedness of sigma-fields, as well as of stochastic variables, is a strict order relation. Note that all propositions in the following theorems are made with respect to a given family $(A_t,\ t \in T)$ of sigma-fields.

3.4. THEOREM

Let the presumptions and notations of 3.2 be valid, let $\mathcal{F} \subset A$ be a sigma-field and let X, W, U be (Ω',A')-, (Ω'',A'')-, and (Ω''',A''')-stochastic variables on the probability space (Ω,A,P), respectively.

(i) If $\mathcal{C}$ is preordered to $\mathcal{D}$, then $\mathcal{D}$ is not preordered to $\mathcal{C}$ (asymmetry).

(ii) If $\mathcal{C}$ is preordered to $\mathcal{D}$ and $\mathcal{D}$ is preordered to $\mathcal{F}$, then $\mathcal{C}$ is also preordered to $\mathcal{F}$ (transitivity).

(iii) Preorderedness of sigma-fields is a strict order relation on the set of all sigma-fields being subsets of A.

(iv) Propositions (i) to (iii) also hold for stochastic variables X, W and U taking the role of $\mathcal{C}$, $\mathcal{D}$, and $\mathcal{F}$, respectively.

Proof.

We only have to show that the properties (i) and (ii) hold for sigma-fields, because a strict order relation is defined by the assymetry and transitivity properties, and preorderedness of stochastic variables is defined through their generated sigma-fields.

(i) If $\mathcal{C}$ is preordered to $\mathcal{D}$ with respect to $(A_t,\ t \in T)$, then there is an element $s \in T$ with $\mathcal{C} \subset A_s$, $\mathcal{D} \not\subset A_s$. As $(A_t, t \in T)$ is monotonically increasing, there is no element $t \in T$, $t > s$, such that $\mathcal{D} \subset A_t$, $\mathcal{C} \not\subset A_t$, which implies that $\mathcal{D}$ is not preordered to $\mathcal{C}$.

(ii) If $\mathcal{C}$ is preordered to $\mathcal{D}$, then

(a) there is a $s \in T$ with $\mathcal{C} \subset A_s$, $\mathcal{D} \not\subset A_s$.

If $\mathcal{D}$ is preordered to F, then

(b) there is a $t \in T$, $t > s$, with $\mathcal{D} \subset A_t$, $F \not\subset A_t$, and

(c) there is an element $u \in T$, $u > t$, with $F \subset A_u$, $F \not\subset A_s$, for all $s \in T$, $s < u$.

Hence, Conditions (a) to (c) imply that there is an element $s \in T$ with $\mathcal{C} \subset A_s$, $F \not\subset A_s$, and there is an element $u \in T$, $u > s$, with $F \subset A_u$, $F \not\subset A_s$, for all $s \in T$, $s < u$.

Theorem 3.4 does not imply that all sigma-fields being subsets of A are in this relation, but that the asymmetry and transitivity conditions are fulfilled.

3.5. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid. If $\mathcal{C}$ is preordered to $\mathcal{D}$, then $\mathcal{C}$ is also preordered to $F := A(\mathcal{C} \cup \mathcal{D})$.

Proof.

If $\mathcal{C}$ is preordered to $\mathcal{D}$, then

(a) there is an element $s \in T$ with $\mathcal{C} \subset A_s$, $\mathcal{D} \not\subset A_s$, and

(b) there is an element $t \in T$, $t > s$, with $\mathcal{D} \subset A_t$, $\mathcal{D} \not\subset A_s$, $\forall\, s \in T$, $s < t$.

$(A_t, t \in T)$ being monotonically increasing, implies

(c) $\mathcal{C} \subset A_t$, $F \subset A_t$, because $\mathcal{C}, \mathcal{D} \subset A_t$ implies $F \subset A_t$.

(d) F is not a subset of A_s, $s \in T$, $s < t$, because of (b).

Hence, there is an element $s \in T$ with $\mathcal{C} \subset A_s$, $F \not\subset A_s$ (see (a) and (d)), and there is a $t \in T$, $t > s$, with $F \subset A_t$, $F \not\subset A_s$, $\forall\, s \in T$, $s < t$ (see (c) and (d)), which proves the theorem.

According to the following theorem, equiorderedness is an equivalence relation.

3.6. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid.

(i) If there is a $t \in T$ with $\mathcal{C} \subset A_t$ and no $s \in T$, $s < t$, with $\mathcal{C} \subset A_s$, then $\mathcal{C}$ is equiordered to itself (reflexivity).

(ii) If $\mathcal{C}$ and $\mathcal{D}$ are equiordered, then $\mathcal{D}$ and $\mathcal{C}$ are equiordered (symmetry).
(iii)If $\mathcal{C}$ and $\mathcal{D}$ as well as $\mathcal{D}$ and $\mathcal{F}$ are equiordered, then $\mathcal{C}$ and $\mathcal{F}$ are equiordered (transitivity).
(iv) Equiorderedness of sigma-fields is an equivalence relation on the set of all sigma-fields $\mathcal{C}_i \subset \mathcal{A}$, for which there is an element $t \in T$ with $\mathcal{C}_i \subset \mathcal{A}_t$ and no $s \in T$, $s < t$, with $\mathcal{C}_i \subset \mathcal{A}_s$.
(v) Propositions (i) to (iv) also hold for equiorderedness of stochastic variables X, W, U taking the roles of $\mathcal{C}$, $\mathcal{D}$, and $\mathcal{F}$, respectively.

Proof.

(i) Conditions 3.3a and 3.3b are obviously true for $\mathcal{C} = \mathcal{D}$.

(ii) Is obvious.

(iii) If $\mathcal{C}$ and $\mathcal{D}$ are equiordered, then

(a) there is a smallest $t \in T$ such that $\mathcal{C}, \mathcal{D} \subset \mathcal{A}_t$ and no $s \in T$, $s < t$, with $\mathcal{C} \subset \mathcal{A}_s$ or $\mathcal{D} \subset \mathcal{A}_s$.

If $\mathcal{D}$ and $\mathcal{F}$ are equiordered, this implies that also

(b) $\mathcal{F} \subset \mathcal{A}_t$ and that

(c) there is no $s \in T$, $s < t$, with $\mathcal{D} \subset \mathcal{A}_s$ or $\mathcal{F} \subset \mathcal{A}_s$.

(a) to (c) imply that there is a $t \in T$ with $\mathcal{C}, \mathcal{F} \subset \mathcal{A}_t$, whereas (a) and (c) imply that there is no $s \in T$, $s < t$, with $\mathcal{C} \subset \mathcal{A}_s$ or $\mathcal{F} \subset \mathcal{A}_s$. Hence, $\mathcal{C}$ and $\mathcal{F}$ are equiordered, which implies the transitivity property.

(iv) An equivalence relation is defined by reflexivity, symmetry, and transitivity.

(v) This proposition is implied by 3.6i to 3.6iv, if we define $\mathcal{C} := X^{-1}(\mathcal{A}')$, $\mathcal{D} := W^{-1}(\mathcal{A}'')$, and $\mathcal{F} := U^{-1}(\mathcal{A}''')$.

It should be noted that not every sigma-field $\mathcal{C} \subset \mathcal{A}$ is equiordered to itself. It may happen that $\mathcal{C}$ is not in the equiorderedness relation, because there is no smallest $t \in T$ with $\mathcal{C} \subset \mathcal{A}_t$. If $T = \mathbb{R}$, $\mathcal{C} = \{\emptyset, \Omega\}$, for example, there is no smallest $t \in T$ with $\{\emptyset, \Omega\} \subset \mathcal{A}_t$. The equivalence relation of equiorderedness is not defined on the set of all sigma-fields $\mathcal{C}_i \subset \mathcal{A}$, but only on the set of all $\mathcal{C}_i \subset \mathcal{A}$, for which there is a smallest $t \in T$ with $\mathcal{C}_i \subset \mathcal{A}_t$.

3.7. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid.

If $\mathcal{C}$ and $\mathcal{D}$ are equiordered, then $\mathcal{C}$ and $\mathcal{F} := \mathcal{A}(\mathcal{C} \cup \mathcal{D})$, as well as $\mathcal{D}$ and

$\mathcal{F}$ are equiordered.

Proof.

If $\mathcal{C}$ and $\mathcal{D}$ are equiordered, then there is a smallest $t \in T$ such that

(a) $\mathcal{C}, \mathcal{D} \subset A_t$, and

(b) there is no $s \in T$, $s < t$, with $\mathcal{C} \subset A_s$ or $\mathcal{D} \subset A_s$.

Condition (a) implies $\mathcal{F} \subset A_t$, because $\mathcal{F}$ is the smallest sigma-field containing both $\mathcal{C}$ and $\mathcal{D}$ as subsets. Condition (b) implies that there is no $s \in T$, $s < t$, with $\mathcal{C}$, $\mathcal{D}$, or $\mathcal{F} \subset A_s$. This implies that $\mathcal{C}$ and $\mathcal{F}$ as well as $\mathcal{D}$ and $\mathcal{F}$ are equiordered.

3.8. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid.

(i) If $\mathcal{C}$ is preordered to $\mathcal{D}$, and $\mathcal{D}$ and $\mathcal{F}$ are equiordered, then $\mathcal{C}$ is preordered to $\mathcal{F}$.

(ii) If $\mathcal{C}$, $\mathcal{D}$ are equiordered and $\mathcal{D}$ is preordered to $\mathcal{F}$, then $\mathcal{C}$ is preordered to $\mathcal{F}$.

(iii) Propositions (i) and (ii) also hold for stochastic variables X, W and U taking the roles of $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{F}$, respectively.

Proof.

(i) In this case,

(a) there is an element $s \in T$ with $\mathcal{C} \subset A_s$, $\mathcal{D} \not\subset A_s$ and an element $t \in T$, $t > s$, with $\mathcal{D} \subset A_t$, $\mathcal{D} \not\subset A_t$, for all $s \in T$, $s < t$.

Furthermore,

(b) $\mathcal{F} \subset A_t$ and there is no $s \in T$, $s < t$, with $\mathcal{D} \subset A_s$ or $\mathcal{F} \subset A_s$, for all $s \in T$, $s < t$.

(a) and (b) imply that there is a $s \in T$ with $\mathcal{C} \subset A_s$, $\mathcal{F} \not\subset A_s$, and an element $t \in T$, $t > s$, with $\mathcal{F} \in A_t$, $\mathcal{F} \not\subset A_s$, for all $s \in T$, $s < t$, which implies that $\mathcal{C}$ is preordered to $\mathcal{F}$ (see 3.2).

(ii) Is obvious.

(iii) This proposition immediately follows from 3.8i and 3.8ii, if we define $\mathcal{C} := X^{-1}(A')$, $\mathcal{D} := W^{-1}(A'')$, and $\mathcal{F} := U^{-1}(A''')$.

Another concept needed in the following sections is that of equi-or preorderedness, where 'or' denotes the logical disjunction. Equi- or preorderedness is a weak order relation. Again, the propositions are always made with respect to an underlying family $(A_t, t \in T)$ of sigma-fields, the explicit reference to which is omitted only for reasons of convenience.

3.9. DEFINITION

Let the presumptions and notations of 3.2 be valid.

(i) We say that C is equi- or preordered to $\mathcal{D}$, iff

(a) C and $\mathcal{D}$ are equiordered, or

(b) C is preordered to $\mathcal{D}$.

(ii) We say that X is equi- or preordered to W, iff C is equi- or preordered to $\mathcal{D}$ and

(c) $C = X^{-1}(A')$ and $\mathcal{D} = W^{-1}(A'')$.

3.10. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid.

(i) If there is a $t \in T$ with $C \subset A_t$ such that there is no $s \in T$, $s < t$, with $C \subset A_t$, then C is equi- or preordered to itself (reflexivity).

(ii) If C is equi- or preordered to $\mathcal{D}$, and $\mathcal{D}$ is equi- or preordered to C, then C and $\mathcal{D}$ are equiordered (antisymmetry).

(iii) If C is equi- or preordered to $\mathcal{D}$, and $\mathcal{D}$ is equi- or preordered to F, then C is equi- or preordered to F (transitivity).

(iv) Equi- or preorderedness of sigma-fields is a weak order relation on the set of all sigma-fields $C_i \subset A$, for which there is a $t \in T$ with $C_i \subset A_t$ such that there is no $s \in T$, $s < t$, with $C \subset A_s$.

(v) Propositions 3.10i to 3.10iv also hold for equiorderedness of stochastic variables X, W, U taking the roles of C, $\mathcal{D}$, and F, respectively.

Proof.

Property 3.10i follows from 3.6i. In order to prove 3.10ii, we consider four cases. First, both C and $\mathcal{D}$ are preordered to one another, which can be excluded, because of 3.4i. Second, C and $\mathcal{D}$ are equiordered. In this case, the proposition 3.10ii is implied by 3.6ii. Third, C is preordered to $\mathcal{D}$ and $\mathcal{D}$ is equiordered to C. This contradicts Condition (b) of Definition 3.3. Fourth, $\mathcal{D}$ is preordered to C, and C is equiordered to $\mathcal{D}$. Again, this case cannot occur, because of 3.3b. The proof of property (iii) follows the same line of arguments as that of property 3.10ii. For the first case that C, $\mathcal{D}$ and $\mathcal{D}$, F are equiordered, the property is implied by 3.6iii. For the second case that C is preordered to C and $\mathcal{D}$ is preordered to F, Property 3.10iii is implied by 3.4iii. For the third case that $\mathcal{C}$ is preordered to $\mathcal{D}$, and $\mathcal{D}$ and $\mathcal{F}$ are equiordered, it is implied by 3.8i, and for the last case that $\mathcal{C}$ and $\mathcal{D}$ are equiordered, whereas $\mathcal{D}$ is preordered to $\mathcal{F}$, it is implied by 3.8ii. (iv) A weak order relation is defined by reflexivity, antisymmetry, and transitivity. (v) This proposition follows from (i) to (iv), if we define $\mathcal{C} := X^{-1}(\mathcal{A}')$, $\mathcal{D} := W^{-1}(\mathcal{A}'')$, and $\mathcal{F} := U^{-1}(\mathcal{A}''')$.

3.11. THEOREM

Let the presumptions and notations of 3.2 and 3.4 be valid. If $\mathcal{C}$ is equi- or preordered to $\mathcal{D}$, then $\mathcal{C}$ is also equi- or preordered to $\mathcal{F} := \mathcal{A}(\mathcal{C} \cup \mathcal{D})$.

Proof.

(a) If $\mathcal{C}$ is preordered to $\mathcal{D}$, then, according to 3.5i, $\mathcal{C}$ is also preordered to $\mathcal{F}$.

(b) If $\mathcal{C}$ and $\mathcal{D}$ are equiordered, then, according to 3.7i, $\mathcal{C}$ and $\mathcal{F}$ are also equiordered. (a) and (b) imply the theorem.

4. POTENTIAL DISTURBING SIGMA-FIELDS AND VARIABLES

By intuition, it is obvious that preorderedness of X to Y is a necessary but not sufficient condition for the stochastic dependence of Y on X to be causal. Two alternative second necessary conditions, the invariance and average conditions have been discussed by Steyer (1984). To state either of them precisely, we need a formal definition of a potential disturbing variable and a potential disturbing sigma-field. The aim is to define variables which are possibly confounded with X and thus make the dependence of Y on X a spurious one.

In the following definition we need the concept of probability measure that is nontrivial with respect to a sigma-field $\mathcal{C}$, which is defined to be a probability measure Q on $\mathcal{C}$ such that there is at least one $C \in \mathcal{C}$ with $0 < Q(C) < 1$.

4.1. DEFINITION

Let $(\Omega,\mathcal{A},P)$ be a probability space, let $(\mathcal{A}_t,\ t \in T)$, $T \subset \mathbb{R}$, be a monotonically increasing family of sigma-fields with $\mathcal{A}(\cup_{t\in T} \mathcal{A}_t) \subset \mathcal{A}$, let $\mathcal{C},\mathcal{D},\mathcal{F} \subset \mathcal{A}$ denote sigma-fields, and finally let X, W be $(\Omega',\mathcal{A}')$-, $(\Omega'',\mathcal{A}'')$-stochastic variables on $(\Omega,\mathcal{A},P)$, respectively.

(i) $\mathcal{D}$ is called potential disturbing sigma-field with respect to $(\mathcal{A}_t, t \in T)$ and $\mathcal{C}$ [or, if $\mathcal{C} = X^{-1}(\mathcal{A}')$, with respect to $(\mathcal{A}_t, t \in T)$ and X], iff the following two conditions hold:

(a) $\mathcal{D}$ is equi- or preordered to $\mathcal{C}$, and

(b) there is a probability measure Q on $\mathcal{A}$ that is nontrivial with respect to both $\mathcal{C}$ and $\mathcal{D}$ such that $\mathcal{C}$ and $\mathcal{D}$ are stochastically independent.

(ii) W is called a potential disturbing variable with respect to $(\mathcal{A}_t, t \in T)$ and $\mathcal{C}$ [or, if $\mathcal{C} = X^{-1}(\mathcal{A}')$, with respect to $(\mathcal{A}_t, t \in T)$ and X], iff Conditions (a) and (b) hold, as well as

(c) $\mathcal{D} = W^{-1}(\mathcal{A}'')$.

Condition 4.1a is chosen in order to exclude mediator variables from the set of potential disturbing variables. Mediator variables, say U, are variables for which X is preordered to U and U is preordered to Y. These variables do not threaten the validity of a causal proposition but may elaborate it. They are important to distinguish between direct and total effects. Condition 4.1b is added to exclude stochastic variables like X, X^2, aX, etc. from the set of potential disturbing variables. It is not possible to construct a probability measure Q on $\mathcal{A}$ that is nontrivial with respect to $\mathcal{A}(X)$ and $\mathcal{A}(X^2)$, for example, such that X and X^2 are stochastically independent with respect to Q. Furthermore, if U is a potential disturbing variable, then $X+U$, $X \cdot U$, for example, are not potential variables.

4.2. EXAMPLE

Consider a study on the effectiveness of a retraining program on the rate of recidivism (for a more detailed description of the example see Steyer (1984)). We might choose

$$\Omega = \{(a_i, b_j, c_k)\colon i,j,k = 0,1\},$$

where (a_i, b_j, c_k) represents one of the eight possible elementary events that a parolee is male or female (a_i), takes part at the retraining program or not (b_j), and commits a crime or not after the retraining (c_k). We choose the sigma-field $\mathcal{A}$ on Ω to be the set of all subsets of Ω, which consists of $\emptyset$, the eight sets $\{(a_i, b_j, c_k)\}$ and all unions of these sets. Furtheron, we might choose the stochastic variables

$$W = \begin{cases} 1, & \text{if the parolee is female} \\ 0, & \text{otherwise} \end{cases}$$

$$X = \begin{cases} 1, & \text{if the parolee takes part at the retraining} \\ 0, & \text{otherwise} \end{cases}$$

$$Y = \begin{cases} 1, & \text{if there is no recidivism for the parolee} \\ 0, & \text{otherwise,} \end{cases}$$

to represent the events that might be observed. If $T := \{1,2,3\}$, $A_1 := A(W)$, $A_2 := A(W,X)$, and $A_3 := A(W,X,Y)$, then W is preordered to both X and Y, and X is preordered to Y. Furthermore, W then is a potential disturbing variable with respect to X and $(A_t,\ t \in T)$, because W is equi- or preordered to X and 4.1b is fulfilled, too, as is easily seen, if we choose $Q(X=x) = Q(W=w) = 0.5$ and $Q(X=x, W=w) = 0.25$, for all combinations of values of X and W.

Obviously, the formalization of this example, consisting of only three dichotomous variables, is very restricted. It represents only a very small part of the reality investigated. It does not represent all variables that might be important for the phenomena of recidivism of parolees. It does not contain variables like the parolee's marital status, the kind of crime committed, being raised or not in a complete family, the parolees educational background, etc. It would be possible, of course, to formulate a new model including all these variables. However, such an explicit enumeration of variables occuring in the part of reality studied, would never be complete. Therefore, an alternative kind of application of the formal theory should be considered.

First, one may interpret the probability space (Ω, A, P) to represent a specified experiment, or more generally, the part of reality to be described. Second, one might define $T := \mathbb{R}$ and interpret each sigma-field A_t, $t \in T$, to represent all events which may occur up to point t of time, inclusively. For this kind of interpretations of the formal structures (Ω, A, P) and $(A_t,\ t \in T)$, it is not necessary to know all the events in each A_t or even enumerate them explicitely. If (Ω, A, P) represents a specified part of reality, there is no ambiguity about the variables and events involved or about their temporal ordering so long as it is made clear enough which part of reality is considered.

If we follow this second path of interpreting the formal theory in the study on the effectiveness of the retraining with parolees, all properties of the parolee before the retraining are potential disturbing variables, e.g. the kind of the parolee's crime, the kind of experiences in jail, etc. Obviously, a causal proposition is in full accordance with our intuition only for this second kind of interpretation of the sigma-fields A_t, allowing for all potential disturbing variables, disturbing variables which are there, even though they might not be observed.

However, so long as one mentions with respect to which family (A_t, $t \in T$) a causal proposition is made, causal propositions as defined in the next section are meaningful and not trivial, so long as at least one potential disturbing variable is allowed for. The more potential disturbing variables are implicitly or explicitly allowed for by the interpretation of the sigma fields A_t, $t \in T$, the more meaningful a causal proposition is. Only if the A_t are interpreted to represent all events that may occur up to point t of time inclusively, in the real process considered, will causal propositions fully coincide with the intuitive meaning of the term 'causal'. It should be noted, however, that the formal theory proposed is neutral with respect to these questions of adequate application.

5. CAUSAL LINEAR STOCHASTIC DEPENDENCE

Using the concepts of pre- and equiorderedness and of potential disturbing variables treated in the previous sections, as well as that of a conditional expectation, we may now formally define causal linear stochastic dependence and study its formal properties. As not only causal linear stochastic dependence, but also independence is of interest, we first define causal linear stochastic in/dependence, which means dependence or independence. Independence additionally requires $E(Y|C)$ as$=E(Y)$, and dependence the negation of it.

5.1. DEFINITION

Let Y be a real-valued stochastic variable on the probability space (Ω,A,P) with finite expectation $E(Y)$, let X be a (Ω',A')-stochastic variable and W be a (Ω'',A'')-stochastic variable on (Ω,A,P), let $C \subset A$ be a sigma-field, let $(A_t,\ t \in T)$ be a monotonically increasing family of sigma-fields with $A(\cup_{t\in T} A_t) \subset A$, and let $E(Y|C,\mathcal{D}) := E[Y|A(C \cup \mathcal{D})]$ denote the conditional expectation of Y given $A(C \cup \mathcal{D})$. We say that Y is causally linearly *stochastically in/dependent on $\mathcal{C}$ with respect to $(A_t,\ t \in T)$*, iff the following conditions hold:

(a) $\mathcal{C}$ is preordered to $A(Y)$ with respect to $(A_t,\ t \in T)$ (*preorderedness*).

(b) For all potential disturbing sigma-fields $\mathcal{D}$ with respect to $(A_t,\ t \in T)$ and $\mathcal{C}$,

$$E(Y|\mathcal{C},\mathcal{D}) \text{ as= } E(Y|\mathcal{C}) + H\circ W \tag{1}$$

holds, where $H\circ W$ denotes the composition of an A''-measurable real-valued function H with a $(\mathcal{D},A'')$-measurable stochastic variable W (*invariance*).

Iff Conditions (a) and (b), as well as

(c) $E(Y|\mathcal{C})$ as= $E(Y)$ (*linear stochastic independence*) hold, we say that Y is causally linearly stochastically *independent* from $\mathcal{D}$ with respect to $(A_t,\ t \in T)$, and *dependent* on $\mathcal{C}$, otherwise.

If Conditions (a) and (b) hold, as well as (d) $\mathcal{C} = X^{-1}(A')$, we also say that *Y is causally linearly stochastically in/dependent on X with respect to $(A_t,\ t \in T)$*.

If no confusion is possible, we may omit the explicit reference to $(A_t,\ t \in T)$.

There are a number of situations in which Equation (1) holds. Suppose for example, that the conditional expectation of the Y given $\mathcal{C}$ and $\mathcal{D}$ is the sum of a composition $F\circ X$ and a composition $H\circ W$,

$$E(Y|\mathcal{C},\mathcal{D}) \text{ as= } F\circ X + H\circ W$$

where F is an A'-measurable and H an A''-measurable real-valued function, whereas X, W are $(\mathcal{C},A')$-, $(\mathcal{D},A'')$-measurable, respectively. This will be referred to as the *additivity* condition, a special case of which is the multiple regression equation

$$E(Y|X,W) \text{ as= } b_{Y0} + b_{YX}X + b_{YW}W.$$

The additivity condition also holds, for example, if

$$E(Y|X,W) \text{ as= } b_{Y0} + b_{YX}X + b_{YX2}X^2 + b_{YW}W + b_{YW2}W^2,$$

whereas it does not hold, for example, if

$$E(Y|X,W) \text{ as= } b_{Y0} + b_{YX}X + b_{YW}W + b_{Y(X\cdot W)}X\cdot W \text{ and } b_{Y(X\cdot W)} \neq 0.$$

This last equation contradicts the additivity condition, because of the multiplicative term $X\cdot W$, which is neither $\mathcal{C}$- nor $\mathcal{D}$-measurable.

5.2. THEOREM

Let the presumptions and notations of 5.1 be valid. If

(a) for all potential disturbing sigma-fields $\mathcal{D}$,

$$E(Y|\mathcal{C},\mathcal{D}) \text{ as}= F\circ X + H\circ W, \tag{2}$$

where F is an A'- and H an A''-measurable real-valued function, X is $(\mathcal{C},A')$- and W is $(\mathcal{D},A'')$-measurable (additivity); if

(b) X and W are stochastically independent, and if

(c) $\mathcal{C}$ is preordered to $A(Y)$ with respect to $(A_t,\ t \in T)$, then Y is causally linearly stochastically in/dependent on $\mathcal{C}$.

Proof.

In order to prove this proposition, we suppose that the expectation $E(H\circ W)$ is equal to zero. Note that this does not restrict generality, because, by subtracting its expectation, any composition $H'\circ W$ with an expectation unequal zero can easily be transformed into a composition $H\circ X$ with expectation zero. It is easily seen that the conditions (a) and (b) of theorem 5.2 imply Equation (1), because

$$\begin{aligned} E(Y|X) &\text{ as}= E[E(Y|\mathcal{C},\mathcal{D})|X] && \text{(see A.17)}\\ &\text{as}= E(F\circ X+H\circ W|X) && \text{(see 2)}\\ &\text{as}= E(F\circ X|X) + E(H\circ W|X) && \text{(see A.15)}\\ &\text{as}= F\circ X + E(H\circ W) = F\circ X\ , && \text{(see A.9)} \end{aligned}$$

where $E(H\circ W) = 0$ and the theorem is used that stochastic independence of W and X implies $E(H\circ W|X)$ as= $E(H\circ W)$. We may now insert $F\circ X$ as= $E(Y|X)$ into Equation (2), which yields Equation (1), which proves the theorem.

We now turn to another situation in which Y is causally linearly stochastically in/dependent on $\mathcal{C}$. Suppose, $E(Y|X,W)$ as= $E(Y|X)$. In this situation, too, Equation (1) is fulfilled. W being a constant is a special case in which $E(Y|X,W)$ as= $E(Y|X)$ is true. Hence, if there is the possibility that a potential disturbing variable is an actual one, it might be held constant at one of its values. Thus, the experimental control technique of holding potential disturbing variables constant, can be based on the invariance condition. It should be noted, however, that the validity of a causal proposition is then restricted to the case that W is constant at its value w.

5.3. THEOREM

Let the presumptions and notations of 5.1 be valid. If
(a) C is preordered to $\mathcal{A}(Y)$ with respect to $(\mathcal{A}_t,\ t \in T)$, and if
(b) for all potential disturbing sigma-fields $\mathcal{D}$,

$$E(Y|C,\mathcal{D}) \text{ as= } E(Y|C), \tag{3}$$

(C-conditional linear stochastic independence of Y from $\mathcal{D}$), then Y is causally linearly stochastically in/dependent on C.

Proof.

Equation (3) implies Equation (1), because we may define H and W such that $H \circ W = 0$.

5.4. COROLLARY

Let the presumptions and notations of 5.1 be valid. If C is preordered to $\mathcal{A}(Y)$ with respect to $(\mathcal{A}_t,\ t \in T)$ and

$$E(Y|C) \text{ as= } Y \tag{4}$$

(complete dependence of Y on C), then Y is causally linearly stochastically in/dependent on C.

Proof.

$E(Y|C)$ as= Y implies
$E(Y|C,\mathcal{D})$ as= $E[E(Y|C)|C,\mathcal{D}]$
as= $E(Y|C)$, for all sigma-fields $\mathcal{D} \subset \mathcal{A}$, (see A.12)
which implies 5.3b.

Hence, causal linear stochastic dependence does neither require all of the variation of the dependent variable Y to be determined by C (see Equation (4)), nor is it necessary that all influencing variables that are equi- or preordered to C are known (see Equation (3)). It is only required that all potential disturbing sigma-fields $\mathcal{D}$ behave as described by Equation (1). This equation is fulfilled already if Y is C-conditionally linearly stochastically independent from $\mathcal{D}$, i.e. if Equation (3) holds, or if X and W are stochastically independent and additivity can be assumed.

6. SIMPLE CAUSAL REG-LINEAR DEPENDENCE

In the case of a simple reg-linear dependence of Y on X which is characterized by the simple regression equation

$$E(Y|X) \text{ as=} a_{Y0} + a_{YX}X, \tag{5}$$

the invariance condition implies that, for all potential disturbing variables W,

$$E(Y|X,W) \text{ as=} a_{Y0} + a_{YX}X + H\circ W. \tag{6}$$

Whatever the type of the function $H\circ W$, the regression coefficient a_{YX} is left unchanged when turning from Equation (5) to (6), provided the invariance condition holds.

Another remarkable property of the simple regression coefficient a_{YX} is that it is equal to the corresponding regression coefficient $a_{YX|W=w}$ of the conditional simple regression of Y on X given W=w:

$$E(Y|X,W=w) \text{ as=} a_{Y0|W=w} + a_{YX|W=w}X,$$

if the invariance condition holds. This is easily seen from Equation (6), because, for W=w,

$$E(Y|X,W=w) \text{ as=} a_{Y0} + a_{YX}X + H(w),$$

where H(w) is a constant so that we may define $a_{Y0|W=w} := a_{Y0} + H(w)$ and $a_{YX|W=w} := a_{YX}$. Thus, Equation (5) and the invariance condition imply, that a_{YX} is also the regression coefficient of the conditional simple regression of Y on X given W=w, where W is any potential disturbing variable. This motivates the following

6.1. DEFINITION

Let the presumptions and notations of 5.1 be valid. We say that Y is simply causally reg-linearly in/dependent on X and that a_{YX} is the simple causal reg-linear effect of X on Y, iff Y is causally linearly stochastically in/dependent on X and Equation (5) holds.

7. WEAK CAUSAL LINEAR STOCHASTIC DEPENDENCE

It may be argued that the kind of causal linear stochastic dependence discussed above is unrealistically restrictive. Let us consider once again the invariance condition saying that, for all potential disturbing sigma-fields $\mathcal{D}$,

$$E(Y|C,\mathcal{D}) \text{ as}= E(Y|C) + H\circ W,$$

where $H\circ W$ denotes the composition of an A''-measurable real-valued function H with a $(\mathcal{D},A'')$-measurable stochastic variable W. If, for example, C is generated by the real-valued $N(J)$-dimensional stochastic variable $X = (X_1,\ldots,X_{N(J)})$, it is easily seen that there are no multiplicative terms involving a $\mathcal{D}$-measurable function and one or more of the random variables X_j, $j \in J = \{1,\ldots,N(J)\}$. Therefore, the invariance condition excludes interactions in the analysis of variance sense of any order between the variables X_j and any potential disturbing variable W. This condition is a rather strict one that cannot be guaranteed to hold even in randomized experiments, where the term 'causal' seems to be appropriate, too, to characterize the dependence of Y on the treatment variables X_j indicating membership to the experimental groups.

Therefore, a weak type of causal linear stochastic dependence is now introduced. All that has to be done is to replace the invariance by the average condition, which can be formulated as follows: for all potential disturbing variables W, the following equation holds for P_X-almost all $E(Y|X=x)$:

$$E(Y|X=x) = \int E(Y|X=x,W=w)\; P_W(dw),$$

where P_W denotes the distribution of W. If W is a discrete random variable, this equation can be written

$$E(Y|X=x) = \Sigma_w E(Y|X=x,W=w) \cdot P(W=w),$$

where the summation is over all values w of W. This average condition simply means that P_X-almost all conditional expected values $E(Y|X=x)$ are the average of the conditional expectations $E(Y|X=x,W=w)$ of Y given $X=x$ and $W=w$ across the values w of W. Together with preorderedness, the average condition defines weak causal linear stochastic dependence of Y on X.

7.1. DEFINITION

Let the presumptions and notations of 5.1 be valid. We say that Y is weakly causally linearly stochastically in/dependent on X, iff

(a) X is preordered to Y with respect to $(A_t,\ t \in T)$ (preorderedness).

(b) For all potential disturbing variables W with respect to X and $(A_t,\ t \in T)$

$$E(Y|X=x) = \int E(Y|X=x,W=w)\ P_W(dw), \tag{7}$$

for P_X-almost all $x \in \Omega'$ (average condition). If additionally

(c) $E(Y|X) \text{ as}= E(Y)$ (linear stochastic independence),

we say that Y is weakly causally linearly stochastically independent from X and, dependent on X, otherwise.

A special situation in which Equation (7) is fulfilled is characterized by $E(Y|X,W) \text{ as}= E(Y|X)$, which defines X-conditional linear stochastic independence of Y from W.

7.2. THEOREM

Let the presumptions and notations of 5.1 be valid. If

(a) X is preordered to Y with respect to $(A_t,\ t \in T)$ and

(b) for all potential disturbing variables W,

$$E(Y|X,W) \text{ as}= E(Y|X), \tag{8}$$

then Y is weakly causally linearly stochastically in/dependent on X.

Proof.

Equation (8) implies $E(Y|X=x,W=w) = E(Y|X=x)$, for $P_{(X,W)}$-almost all $(x,w) \in \Omega' \times \Omega''$. Hence,

$\int E(Y|X=x,W=w)\ P_W(dw) = \int E(Y|X=x)\ P_W(dw) = E(Y|X=x)$.

The importance of the average condition, however, stems from the fact that it is implied by the stochastic independence of all potential disturbing variables W at one side and the N(J)-dimensional variable $X = (X_1,\dots,X_{N(J)})$ on the other side.

7.3. THEOREM

Let the presumptions and notations of 5.1 be valid. If

(a) X is preordered to Y and
(b) X and all potential disturbing variables are stochastically independent,
then Y is weakly causally linearly stochastically in/dependent on X.

Proof.

A.17 yields $E[E(Y|X,W)|X] \overset{as}{=} E(Y|X)$, which implies

(a) $E(Y|X=x) = E[E(Y|X,W)|X=x]$, for P_X-almost all $x \in \Omega'$.

Stochastic independence of X and W then implies

$E(Y|X=x) = E[E(Y|X=x,W)] = \int E(Y|X=x,W=w)\, P_W(dw)$

where the last line follows from (a) and 5.3.22 of Gänssler and Stute ((1977), p. 169).

In a randomized experiment, $X = (X_1,\dots,X_{N(J)})$ may represent the vector of treatment variables indicating the experimental group to which an experimental unit (subject) belongs. *Randomization* implies that all potential disturbing variables W (representing any properties of the experimental units before or at the treatment) and the treatment variable $X = (X_1,\dots,X_{N(J)})$ are stochastically independent. Therefore, the average condition holds in randomized experiments. Thus, weak causal linear stochastic dependence (or independence, if $E(Y|X) \overset{as}{=} E(Y)$) is guaranteed by randomization (for an example see Steyer (1984)). *Matching*, of course, serves the same purpose, but only for those potential disturbing variables W with respect to which the experimental groups are matched. Thus, the control techniques of randomization and matching can be based on the theory proposed, too.

However, and maybe more importantly, the theory of causal linear stochastic dependence is also relevant for *nonexperimental studies*, because both the *invariance* and the *average condition* may be tested, once a potential disturbing variable W is observed. If neither holds, any causal hypothesis should be rejected for the model considered. If one or both of them hold, one may maintain the causal hypothesis (in its weak or its strict form) as long as no potential disturbing variable is found for which the invariance and the average conditions do not hold. Testing if the average condition (see Equation (7)) holds is a test of weak causal linear stochastic dependence and testing if the invariance condition (see Equation (2)) holds is a test of causal linear stochastic dependence (in the stricter sense). Of course, such tests make sense only if X is assumed to be preordered to Y.

In the last theorem, it is shown that causal linear stochastic in/dependence implies weak causal linear stochastic in/dependence.

7.4. THEOREM

Let the presumptions and notations of 5.1 be valid. If Y is causally linearly stochastically dependent on X, then Y is also weakly causally linearly stochastically dependent on X.

Proof.

We have to show that Equation (1) implies Equation (7). Equation (1) implies:

$$E(Y \mid X=x, W=w) = E[E(Y \mid X, W) \mid X=x, W=w]$$

$$= E[E(Y \mid X) \mid X=x, W=w] + E(H \circ W \mid X=x, W=w) = E(Y \mid X=x) + H(w).$$

Hence, $E(Y \mid X=x) = E(Y \mid X=x, W=w) - H(w)$ and

$$E(Y \mid X=x) = \int E(Y \mid X=x)\, P_W(dw) = \int [E(Y \mid X=x, W=w) - H(w)]\, P_W(dw)$$

$$= \int E(Y \mid X=x, W=w)\, P_W(dw) - \int H(w)\, P_W(dw) = \int E(Y \mid X=x, W=w)\, P_W(dw),$$

because $\int H(w)\, P_W(dw) = E(H \circ W) = 0$, which is easily seen from

$$E[E(Y \mid X, W)] = E[E(Y \mid X)] + E(H \circ W),$$

which implies $E(H \circ W) = 0$, because $E[E(Y \mid X, W)] = E[E(Y \mid X)] = E(Y)$.

For examples illustrating the application of the theory proposed, as well as for a discussion relating this theory to simultaneous equation modeling, the reader is referred to Steyer (1984).

APPENDIX

In this appendix are presented the definition of, and some theorems pertinent to the conditional expectation $E(Y \mid C)$ of Y given C that are frequently cited in this chapter. For an introduction into the concept and its background see e.g. Bauer (1974), Breiman (1968), Gännsler and Stute (1977), or Loève (1977, 1978). A shorter and therefore more convenient name for $E(Y \mid C)$ is *C-conditional expectation of Y*. The C-conditional expectation of Y is a very general and useful concept. It is used, for example, to define the C-conditional probability $P(A \mid C)$ of an event A, as well as the C-conditional variance $V(Y \mid C)$ and covariance $C(Y,Z \mid C)$ of stochastic variables. Special C-conditional expectations are also obtained, if C is the sigma-field generated by a stochastic variable X or by a family $(X_j,\ j \in J)$ of stochastic variables. In these cases, the notations $E(Y \mid X)$, $E(Y \mid X_j,\ j \in J)$, etc. may be used.

The mathematical definition given below does not appeal very much to intuitive insight. Therefore, it might be helpful to recall that $E(Y|X)$ is a stochastic variable, the values of which are identical with the conditional expectations $E(Y|X=x)$ given $X=x$. Another way to think about $E(Y|X)$ is that it is a stochastic variable consisting of the best predictions of Y given a value x of X. In many textbooks on applied statistics, the variable consisting of the best predictions of Y is denoted by $\hat{Y}$. $E(Y|X)$ may be thought of as the mathematical equivalent of $\hat{Y}$. Similarly, $E(Y|\mathcal{C})$ consists of the best predictions of Y based on the sigma-field $\mathcal{C}$.

Definition 1

Let $Y: \Omega \to \mathbb{R}$ be a real-valued stochastic variable on the probability space $(\Omega,\mathcal{A},P)$ with finite expectation $E(Y)$ and let $\mathcal{C} \subset \mathcal{A}$ be a sigma-field. The stochastic variable $E(Y|\mathcal{C}): \Omega \to \mathbb{R}$ is called the $\mathcal{C}$-conditional expectation of Y, iff $E(Y|\mathcal{C})$ is $\mathcal{C}$-measurable and

$$E[1_C E(Y|\mathcal{C})] = E(1_C Y), \quad \text{for all } C \in \mathcal{C}. \tag{1}$$

Let $A \in \mathcal{A}$ be an event and 1_A its indicator function. $P(A|\mathcal{C}): \Omega \to \mathbb{R}$ is called the $\mathcal{C}$-conditional probability of A, iff

$$P(A|\mathcal{C}) := E(1_A|\mathcal{C}). \tag{2}$$

Let X be a $(\Omega',\mathcal{A}')$-stochastic variable on $(\Omega,\mathcal{A},P)$. $E(Y|X)$ is called the X-conditional expectation of Y, iff $\mathcal{C} = \mathcal{A}(X,\mathcal{A}')$ is the sigma-field generated by X and $\mathcal{A}'$, and

$$E(Y|X) := E(Y|\mathcal{C}). \tag{3}$$

Let $(X_j,\ j \in J)$ be a sequence of $(\Omega_j,\mathcal{A}_j)$-stochastic variables on $(\Omega,\mathcal{A},P)$. $E(Y|X_j,\ j \in J)$ is called the $(X_j,\ j \in J)$-conditional expectation of Y, iff $\mathcal{C} = \mathcal{A}[\cup_{j\in J} \mathcal{A}(X_j,\mathcal{A}_j)]$ is the sigma-field generated by $[(X_j,\mathcal{A}_j),\ j \in J]$, and

$$E(Y|X_j,\ j \in J) := E(Y|\mathcal{C}). \tag{4}$$

$E(Y|X_1,\ldots,X_{N(J)})$ is called the $(X_1,\ldots,X_{N(J)})$-conditional expectation of Y, iff $\mathcal{C} = \mathcal{A}[\cup_{j\in J} \mathcal{A}(X_j,\mathcal{A}_j)]$ is the sigma-field generated by $[(X_j,\mathcal{A}_j),\ j \in J]$, where $J = \{1,\ldots,N(J)\}$, $N(J) \in \mathbb{N} = \{1,2,\ldots\}$, and

$$E(Y|X_1,\ldots,X_{N(J)}) := E(Y|\mathcal{C}). \tag{5}$$

Note that $E(Y|C)$ is the general concept. Hence, if propositions are true for $E(Y|C)$, then they also hold for $P(A|C)$, $E(Y|X)$, $E(Y|X_j, j \in J)$, and $E(Y|X_1,\ldots,X_{N(J)})$. Through the definition above, the C-conditional expectation $E(Y|C)$ is uniquely determined only with probability one. Therefore, there are generally different versions of $E(Y|C)$, which are, however, equivalent, with probability 1. Equations on $E(Y|C)$ are therefore true only almost surely in general, which will be abbreviated by the symbol 'as='.

Notice that there is a difference between the C-conditional expectation $E(Y|C)$ of Y, i.e. the conditional expectation of Y *with respect to the sigma-field C* and the conditional expectation $E(Y|C)$ of Y *given the event C*, and correspondingly, a difference between the conditional probability $P(A|C)$ of the event A *with respect to the sigma-field* C and the conditional probability $P(A|C)$ of the event A *given the event C*. $E(Y|C)$ and $P(A|C)$ are stochastic variables, whereas $E(Y|C)$ and $P(A|C)$ are real-valued constants.

In the following theorem, a number of propositions on $E(Y|C)$ are gathered. Some of them are on conditions under which $E(Y|C)$ is equal to a constant function. In these and related contexts, the symbol for the function and the constant (its values) will be the same. The symbol 'a', for example, does not only denote a real constant, but also a function $a\colon \Omega \to \mathbb{R}$ taking the value a for all $\omega \in \Omega$. The notation $E(Y|C) \geq 0$, almost sure, in Proposition (11) is an abbreviation for $E(Y|C)(\omega) \geq 0(\omega)$, for almost all $\omega \in \Omega$, where $0\colon \Omega \to \mathbb{R}$ is a function defined by $0(\omega) = 0$, for all $\omega \in \Omega$.

THEOREM 1

Let Y be a real-valued stochastic variable on the probability space (Ω,A,P) with finite expectation $E(Y)$. If $E(Y|C)$ is the C-conditional expectation of Y, then the following propositions are true:

$$E[E(Y|C)] = E(Y) \tag{6}$$

$$E(Y|C) = E(Y), \text{ if } C = \{\Omega,\emptyset\}, \text{ where } E(Y)\colon \Omega \to \mathbb{R} \text{ is defined by } E(Y)(\omega) = E(Y), \text{ for all } \omega \in \Omega \tag{7}$$

$$E[Y-E(Y|C)] = 0. \tag{8}$$

$$E(Y|C) \text{ as= } Y, \text{ if } Y \text{ is } C\text{-measurable} \tag{9}$$

$$E(Y|C) \text{ as= } a, \text{ if } Y \text{ as= } a, \text{ where } a\colon \Omega \to \mathbb{R} \text{ is defined by } a(\omega) = a, \text{ for all } \omega \in \Omega \tag{10}$$

$$E(Y|C) \geq 0, \text{ almost sure, if } Y \geq 0, \text{ almost sure, where } 0\colon \Omega \to \mathbb{R} \text{ is defined by } 0(\omega) = 0, \text{ for all } \omega \in \Omega\,. \tag{11}$$

According to Equation (6), the expectation of the C-conditional expectation of Y is equal to the expectation of Y. Proposition (7) shows that the C-conditional expectation of Y is equal to the constant function $E(Y): \Omega \to \mathbb{R}$, if $C = \{\Omega, \emptyset\}$. According to Equation (8), the expectation of the difference $Y-E(Y|C)$ of Y and its C-conditional expectation is zero. Note that the difference $Y-E(Y|C)$, called the error or residual of Y with respect to $E(Y|C)$, plays an important role in applications. Proposition (9) is often applied. Y is C-measurable, for example, if C is the sigma-field generated by the (Ω_j, A_j)-stochastic variables X_j, $j \in J$, and $Y = X_i$, $Y = X_j \cdot X_k$, or $Y = X_j + X_k$, where $j, k \in J$. Hence, special cases of Proposition (9) are, for example,

$$E(X|X) \text{ as}= X,$$
$$E(X_i|X_j,\ j \in J) \text{ as}= X_i, \text{ if } i \in J,$$
$$E(X_i \cdot X_k|X_j,\ j \in J) \text{ as}= X_i \cdot X_k, \text{ if } i,k \in J,$$
$$E(X_i + X_k|X_j,\ j \in J) \text{ as}= X_i + X_k, \text{ if } i,k \in J.$$

The following theorem contains propositions on $E(Y|C)$, where Y is the weighted sum or the product of other stochastic variables. Again, the notation $Y \leq Z$, almost sure, which occurs in Proposition (16), is an abbreviation for $Y(\omega) \leq Z(\omega)$, for almost all $\omega \in \Omega$.

THEOREM 2

Let Y and Z be real-valued stochastic variables on the probability space (Ω, A, P) with finite expectations $E(Y)$ and $E(Z)$, respectively, and let C, $C_0 \subset A$ be two sigma-fields. If $E(Y|C)$ and $E(Z|C)$ are the C-conditional expectations of Y and Z, respectively, then the following propositions are true:

$$E(Y|C) \text{ as}= E(Z|C), \text{ if } Y \text{ as}= Z. \tag{12}$$
$$E(Y \cdot Z|C) \text{ as}= Y \cdot E(Z|C), \text{ if } Y \text{ is } C\text{-measurable and if } E(Y \cdot Z) \text{ is finite} \tag{13}$$
$$E[E(Y|C_0) \cdot Z|C] \text{ as}= E(Y|C_0) \cdot E(Z|C), \text{ if } C_0 \subset C. \tag{14}$$
$$E(aY+bZ|C) \text{ as}= aE(Y|C) + bE(Z|C), \text{ if } a, b \in \mathbb{R}. \tag{15}$$
$$E(Y|C) \leq E(Z|C), \text{ almost sure, if } Y \leq Z, \text{ almost sure.} \tag{16}$$

THEOREM 3

Let Y be a real-valued stochastic variable on the probability space (Ω, A, P) with finite expectation $E(Y)$. If $E(Y|C)$ is the C-conditional expectation of Y and $C_0 \subset C$ is a sigma-field, then the following equations are true:

$$E[E(Y|C)|C_0] \overset{as}{=} E(Y|C_0) \overset{as}{=} E[E(Y|C_0)|C] , \qquad (17)$$

$$E[Y-E(Y|C)|C_0] \overset{as}{=} 0. \qquad (18)$$

The sigma-field C_0 will be a subset of the sigma-field C, for example, if C is generated by the (Ω_j, A_j)-stochastic variables X_j, $j \in J$, and C_0 by the (Ω_k, A_k)-stochastic variables X_k, $k \in K$, where $K \subset J$. Hence, special cases of the Equations (17) are

$$E[E(Y|X_1,X_2) \mid X_1] \overset{as}{=} E(Y|X_1) \overset{as}{=} E[E(Y|X_1)|X_1,X_2] ,$$

$$E[E(Y|X_j,\ j \in J)|X_k,\ k \in K] \overset{as}{=} E(Y|X_k,\ k \in K)$$

$$\overset{as}{=} E[E(Y|X_k,\ k \in K)|X_j,\ j \in J] .$$

Equation (18) reveals that the C_0-conditional expectation of the error or residual $F = Y - E(Y|C)$ is zero, if $C_0 \subset C$. This proposition on F is much stricter than that of Equation (8), according to which the (unconditional) expectation of F is zero.

REFERENCES

[1] Bagozzi, R.P., Causal models in marketing, New York, Wiley (1980).

[2] Bauer, H., Wahrscheinlichkeitstheorie und Grundzüge der Masstheorie, Berlin, de Gruyter (1974).

[3] Breiman, L., Probability, Reading, Mass., Addison-Wesley (1968).

[4] Cook, T.D. & Campbell, D.T., Quasi-experimentation. Design and analysis issues for field settings, Boston, Houghton Mifflin (1979).

[5] Gänssler, P. & Stute, H., Wahrscheinlichkeitstheorie, Berlin, Springer (1977).

[6] Halmos, P.R., Measure theory, Princeton, N.J., Van Nostrand (1969).

[7] Heise, D.R., Causal analysis, New York, Wiley (1975).

[8] Loève, M., Probability theory I, Berlin, Springer (1977).

[9] Loève, M., Probability theory II, Berlin, Springer (1978).

[10] Simon, H.A., On the definition of the causal relation, The Journal of Philosophy 49 (1952), 517-528.

[11] Stegmüller, W., Erklärung, Begründung, Kausalität. Probleme und Resultate der Wissenschaftstheorie und analytischen Philosophie, Band I, Berlin, Springer (1983).

[12] Steyer, R., Structural equations, stability, and equilibrium points in multivariate autoregressive processes. Paper read at the 13th European Mathematical Psychology Group Meeting, Bielefeld, September (1982).

[13] Steyer, R, Modelle zur kausalen Erklärung statistischer Zusammenhänge. In J. Bredenkamp & H. Feger (Eds.), Enzyklopädie der Psychologie, Serie Forschungsmethoden der Psychologie, Band 4, Göttingen, Hogrefe (1983).

[14] Steyer, R., Causal linear stochastic dependencies: An introduction, in J. R. Nesselroade & A. von Eye (Eds.), Individual development and social change: explanatory analysis, New York, Academic Press (1984).

[15] Suppes, P., A probabilistic theory of causality, Amsterdam, North Holland (1970).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

FACTORIZATION AND ADDITIVE DECOMPOSITION OF A WEAK ORDER

Reinhard Suck
Universität Osnabrück

Possibilities of deriving a Cartesian product structure from a weak order are investigated. As an intermediate step a "compensatory structure" is introduced and a representation theorem for this special type proved. Conditions are formulated that guarantee the desired factorization and yield an additive representation. The paper closes with some remarks on possible applications.

1. INTRODUCTION

Conjoint measurement can be thought of as a device for measuring empirical entities entailing a weak order, for which no natural form of concatenation of the objects is defined, but which, nevertheless, should figure a reasonable type of measurement scale, i.e. an interval scale at least.

Denoting the set of objects by A and the weak order by $\lesssim$ the only other primitive in a conjoint measurement system is, that A possesses a product structure, that is to say: there are two sets X_1, X_2 such that

$$A = X_1 \times X_2. \tag{1.1}$$

Thus, the objects in question are in fact pairs of some kind of attributes. In the presence of certain conditions one can show that a representation exists, i.e. a real valued function f, such that both components of a pair contribute additively to its overall measure. To be more precise: for all a, b $\in$ A we have

$$a \lesssim b \quad \text{iff} \quad f(a) \leq f(b) \tag{1.2}$$

and there are two real valued functions f_1 and f_2 on X_1 and X_2 respectively such that for a = (x,y),

$$f(a) = f_1(x) + f_2(y) \qquad (1.3)$$

Examples of empirically relevant conjoint measurement structures are easily provided, see Krantz, Luce, Suppes, Tversky, Chap. 6 (1971).

But there are also plenty of instances, where either the naturally defined components do not yield a conjoint measurement structure, because the other conditions are not fulfilled, or the empirical objects do not comprise a Cartesian product. In both cases one would like to know whether it is possible to find sets X_1, X_2 such that (1.1) holds and $(X_1, X_2, \leq)$ is a conjoint measurement structure. We note in passing that in this case the identity sign in (1.1) should be taken with a grain of salt; as long as (1.1) holds with some kind of isomorphism argument involved, we are still satisfied.

The present paper deals with conditions that yield the existence of sets X_1, X_2 fulfilling (1.1) and a representation of $(A, \leq)$, which exhibits the properties of an additive conjoint structure, i.e. (1.2) and (1.3).

It has enormous consequences to drop the condition that the set of objects consists of pairs, since almost all of the usual axioms of conjoint structures rely heavily on the Cartesian product property.

To approach this objective we first have a closer look at the structural implications this condition has on the weak order $(A, \lesssim)$, when A satisfies (1.1) and $(X_1, X_2, \lesssim)$ is a symmetric additive conjoint measurement structure. Then we proceed to formulate axioms, from which the factorization and a suitable representation theorem can be derived. The paper closes with some remarks on applications of this theory.

The problem bears some resemblance to multidimensional scaling or even factor analysis, because we are virtually extracting components or dimensions from A, which - when properly scaled - account for the empirical findings. The main difference, however, is that we do not suggest another algorithm for fitting data to an additive model but rather assume an axiomatic approach. Furthermore, we are not specifically dealing with proximities or correlation matrices, although this might be of minor importance.

2. PRELIMINARIES

Given a set A and a binary relation R on A. We often use a R b instead of $(a,b) \in R$. If not otherwise stated, relations are always binary.

Let B be a subset of A and R a relation on A. $R/_B$ is the restriction of R to B i.e.

$$R/_B = \{(a,b) \in R;\ a \in B,\ b \in B\}.$$

The set of ordered pairs of two sets A, B is the Cartesian product $A \times B$. Often, we use a b as abbreviation of (a,b), f: $A \to B$ denotes a function with domain A and range f(A) contained in B. f is called surjective, if f(A) = B. The real numbers are denoted by $\mathbb{R}$ and a function f: $A \to \mathbb{R}$ is called real valued. A quasiorder is a reflexive and transitive relation, a weak order is connected (i.e. for all a, b $\in$ A $a \le b$ or $b \le a$) and transitive. Weak orders are usually denoted by $\le$. We understand $\sim$ to be the symmetric part of $\le$ and $<$ the asymmetric part. It is readily shown that the symmetric part of a quasiorder (and a forteriori of a weak order) is an equivalence relation (reflexive, symmetric, transitive). If L is a quasiorder, its symmetric part is denoted by $\tilde{L}$.

Since we deal with several order relations on the same set (and at the same time), we have to distinguish the different equivalence classes. From $\le$ derives the equivalence $\sim$ and the classes $\bar{a}$, from L derives $\tilde{L}$ and a_L. In most cases we use the more precise notation: "$(A,\le)$ is a weak order" for "$\le$ is a weak order"; the same applies for a quasiorder and any other relation.

In a weak order $(A, \le)$ a subset B of A is called order dense, if for all a, a' $\in$ A such that $a < a'$ an element $b \in B$ exists satisfying $a \le b \le a'$. This notion is important in the representation theorem for ordinal measurement, see for example Roberts, p. 111 (1979). Now, we define an additive conjoint structure, using the notion in the sense of Krantz et al., p. 256 (1971).

Let X_1, X_2 be nonempty sets. $(X_1, X_2, \le)$ is an additive conjoint structure iff the following conditions are satisfied for all $x, x', x_1, x_2, x_3 \in X_1$ and $y, y', y_1, y_2, y_3 \in X_2$.

(1) Weak order: $\le$ is a weak order on $X_1 \times X_2$.

(2) Independence: $xy \le x'y$ implies $xz \le x'z$ for all $z \in X_2$ (the other component likewise).

(3) Thomsen condition: if $x_1y_1 \sim x_2y_2$ and $x_2y_3 \sim x_3y_1$ then $x_1y_3 \sim x_3y_2$.

(4) Restricted solvability: if $x_0y_1 \le xy \le x_2y_1$, then there exists $x_1 \in X_1$ such that $xy \sim x_1y_1$ (the other component likewise).

(5) Archimedean Axiom: every strictly bounded standard sequence is finite

(a sequence $(x_i), x_i \in X_1$, $i=1,2,\ldots$ is a standard sequence on X_1 iff there exists $y, y' \in X_2$ such that $x_i y \sim x_{i+1} y'$ for $i=1,2,\ldots$ and y and y' are not indifferent. Similarly a standard sequence on X_2 is defined).

(6) Essentialness: each component is essential i.e. there exists $x, x' \in X_1$ and $y \in X_2$ such that not $xy \sim x'y$ (the other component likewise).

Of course "iff" is short for if and only if.

3. BASIC RESULTS ON SYMMETRIC CONJOINT STRUCTURES

In this paragraph we introduce some more or less obvious properties of an additive conjoint measurement structure. We restrict our attention to symmetric structures for reasons best expounded by scanning Holman's (1971) proof of the representation theorem for additive conjoint structures. Furthermore, we assume the existence of an element $\underline{x}\,\underline{y}$ such that for all $xy \in X_1 \times X_2, \underline{x} \le_1 x$ and $\underline{y} \le_2 y$. Then, two functions $\xi_i: X_1 \times X_2 \to X_i$ are defined:

$\xi_1(x,y)$ is the solution of $xy \sim \xi_1\, \underline{y}$ and $\xi_2(x,y)$ the solution of $xy \sim \underline{x}\, \xi_2$ provided the equations are solvable. With the help of the solvability axiom we could specify conditions that render these equations solvable, but we do not bother to do so, for we are not striving for the most general treatment. Whenever we use $\xi_i(x,y)$ in this chapter, we tacitly assume its existence.

The purpose of the following results is to motivate the conditions chosen in the next chapter to obtain a factorization of $(A, \le)$. To begin with we note that $\le$ being an independent order relation on $X_1 \times X_2$ gives rise to weak orderings $\le_1$ and $\le_2$ on X_1 and X_2 respectively by the definition: for $x,y \in X_1$

$$x \le_1 y \quad \text{iff for some } z \in X_2, xz \le yz$$

($\le_2$ is defined on X_2 similarly).

Each of these relations can be used to develop a new weak order on $X_1 \times X_2$. We define for all $xy, x'y' \in X_1 \times X_2$

$$\begin{aligned} xy \; L \; x'y' \quad &\text{iff } x \le_1 x' \\ xy \; L' \; x'y' \quad &\text{iff } y \le_2 y'. \end{aligned}$$

It is easily confirmed that L and L' are weak orders on $X_1 \times X_2$.

The chief point of interest about L and L' is that they distinguish between elements from the same equivalence class of $\sim$, i.e. if $xy \sim x'y'$ then xy and $x'y'$ are generally not indifferent with respect to L or L'.

Without loss of generality we may simplify things a bit by assuming the equivalence classes of the indifference relation associated with $\leq_i$ to consist of one element only. Denoting indifference with respect to L(L') by $\tilde{L}(\tilde{L}')$ we easily conclude that $xy \sim x'y'$ and $xy \, \tilde{L} \, x'y'$ implies $xy = x'y'$ (the same holds for L'). To verify this assertion assume $xy \sim x'y'$ and $xy \, \tilde{L} \, x'y'$, which implies $x \leq_1 x'$ and $x' \leq_1 x$, so $x \sim_1 x'$. In virtue of the above simplification we have $x=x'$. So $xy \sim xy'$ or $y \leq_2 y'$ and $y' \leq_2 y$, which again yields $y=y'$.

Now, we pick a fixed element from $X_1 \times X_2$, $x_0 y_0$ say, and restrict L to its equivalence class $\overline{x_0y_0}$. This restriction, again denoted by L, is a simple order on $\overline{x_0y_0}$. But a more thorough investigation reveals much more structure of $(\overline{x_0y_0}, L)$, which is due to the other conditions fulfilled by $(X_1,X_2,\leq)$. Take, for example, two elements xy, $x_1y_1 \in \overline{x_0y_0}$ such that $xy \, L \, x_1y_1$. Keeping in mind the additive representation and the indifference of xy and x_1y_1, we know $f_1(x) + f_2(y) = f_1(x_1) + f_2(y_1)$ and $f_1(x) \leq f_1(x_1)$. Thus, $f_2(y) \geq f_2(y_1)$.

Loosely speaking in xy and x_1y_1 both components contribute differently to the overall value of xy and x_1y_1 respectively. What is deficient in one component is compensated for in the other.

In chapter 4 we are going to define a compensatory structure without alluding to components. To this purpose we point out that for given $xy \in \overline{x_0y_0}$ the presupposed existence of $x' = \xi_1(\underline{x},y')$ and $y' = \xi_2(x,\underline{y})$ provides a specific element in $\overline{x_0y_0}$, which in a sense is complementary to xy. First of all, we assert $x'y' \in \overline{x_0y_0}$. This is, in fact, true by the Thomsen condition applied to the equations $x\underline{y} \sim \underline{x}y'$ and $\underline{x}y \sim x'\underline{y}$. It seems to be reasonable to conceive of the element $x'y'$ just defined as a relation on $\overline{x_0y_0}$, which in the context of the component free treatment of the next chapter features as a primitive.
We define

$$xy \; K \; x'y' \text{ iff } x' = \xi_1(\underline{x},y), y' = \xi_2(x,\underline{y}).$$

K enjoys the following properties

$$xy\ K\ x'y' \text{ implies } x'y'\ K\ xy \tag{3.1}$$

$$\text{If } xy \sim x_1y_1,\ xy\ L\ x_1y_1,\ xy\ K\ x'y', \text{ and } x_1y_1\ K\ x_1'y_1', \text{ then } x_1'y_1'\ L\ x'y' \tag{3.2}$$

(3.1) is an easy consequence of the definition of ξ_i. To prove (3.2) we observe that $xy\ L\ x_1y_1$ implies $x \leq_1 x_1$, which together with $xy \sim x_1y_1$ yields $y_1 \leq_2 y$ and $\underline{x}y_1 \leq \underline{x}y$. From this and $\underline{x}y_1 \sim \xi_1(\underline{x},y_1)\ \underline{y}$ and $\underline{x}y \sim \xi_1(\underline{x},y)\ \underline{y}$ we infer $\xi_1(\underline{x},y_1) \leq_1 \xi_1(\underline{x},y)$ or $x_1' \leq_1 x'$. So, by the definition of L, $x_1'y_1'\ L\ x'y'$.

The following property, although not used in the sequel, sheds additional light on the nature of the relation K

If $xy\ K\ x'y'$ and $x'y'\ K\ x''y''$, then $xy\ \tilde{L}\ x''y''$. In the present case $xy\ \tilde{L}\ x''y''$ can be replaced by $xy = x''y''$, since $xy \sim x''y''$.

So far we tried to uncover some properties of the indifference curves of an additive conjoint structure, thereby limiting our attention to one equivalence class only. Next, we characterize L' by L and K. We have

$$xy\ L'\ x_1y_1 \text{ iff } xy\ K\ x'y',\ x_1y_1\ K\ x_1'y_1' \text{ and } x'y'\ L\ x_1'y_1' \tag{3.3}$$

This is proved by the definition of K and the following chain of equivalences:

$xy\ L'\ x_1y_1$ iff $y \leq_2 y_1$, iff $\xi_1(\underline{x},y) \leq_1 \xi_2(\underline{x},y_1)$, iff $x' \leq_1 x_1'$,
iff $x'y'\ L\ x_1'y_1'$.

Now, we deal with a condition concerning the compatibility of $\leq$, L, and K on $X_1 \times X_2$. We exploit the simple fact of additive conjoint structures that an element xy, which is superior to another element x_1y_1 in both components separately, exceeds it with respect to $\leq$, too. In view of the component free treatment we have in mind, we must not use such a description, but convert it to a statement involving L and K only. The condition reads as follows:

$$\text{if } x_1y_1\ L\ xy,\ x_1y_1\ K\ x_1'y_1',\ xy\ K\ x'y', \text{ and } x_1'y_1'\ L\ x'y', \text{ then } x_1y_1 \leq xy \tag{3.4}$$

It should be noted that (3.2) does not render the assumption of (3.4) empty (not even almost empty), because (3.2) assumes $xy \sim x_1y_1$.

The proof of (3.4) is simple, if we observe that by (3.3) we have $x_1y_1\ L'\ xy$. Thus $y_1 \leq_2 y$, but $x_1 \leq_1 x$ follows from $x_1y_1\ L\ xy$. In the next paragraph the relations K and L and its properties (3.1)-(3.4) crop up in a different setting.

4. CONJOINT STRUCTURES WITHOUT COMPONENTS

In this part of the paper we present conditions, from which a factorization and additive representation of the kind detailed in the introduction can be derived. The properties of conjoint structures emphasized in the preceding chapter are crucial for this enterprise.

One of the lessons we learned from it was the compensatory structure of the indifference curves. So, we set about defining, what this is going to be.

DEFINITION

A triple (A, L, K), is a compensatory structure iff A is a non empty set and L and K binary relations satisfying the following axioms.

(1) (A, L) is a weak order
(2) For all $a \in A$ exists $a' \in A$ such that a K a'
(3) K is symmetric
(4) If a L b, a K a', b K b', then b' L a'
(5) If a K a' and a' $\tilde{L}$ b, then a K b.

The real numbers with the usual $\leq$ relation and $K = \{(x,y);\ x+y=1\}$ provide an example of a compensatory structure.

A compensatory structure is another case of fundamental measurement (for a comprehensive discussion of this concept see Roberts (1979)), very similar to ordinal structures, but somewhat enriched due to the relation K. To give an empirical example: in studying auditory perception, sounds of different intensity are presented to the left ear (say). The subject is asked to adjust a second sound, which can be presented to the right ear, in such a manner that the combined loudness of both sounds matches with a fixed reference sound. A suitable representation theorem reads as follows.

THEOREM 1

Let (A, L, K) be a compensatory structure containing an order-dense countable subset. Then, there is a function $f: A \to \mathbb{R}$ satisfying

$$a\ L\ b \text{ iff } f(a) \leq f(b) \qquad (4.1)$$

If a K a' and b K b', then

$$f(a) + f(a') = f(b) + f(b'). \qquad (4.2)$$

Before entering the details of the proof, we sketch its key idea.

Obviously (4.1) offers no difficulty, if it were not for (4.2). So, we first show that it is possible to halve A, thus getting two essentially disjunctive subsets, one consisting of its "small" elements, the other containing the "larger" objects. The latter can be shown to coincide with the compensating elements of the former set. Consequently, an arbitrary ordinal representation of the "small" set, can be transformed and extended to the "large" set in such a way that (4.1) and (4.2) hold simultaneously. The complete proof is split into a series of lemmas. The reader not particularly interested may skip this part and resume reading with Theorem 2.

LEMMA 1: $a K b$ and $b K c$ implies $a \tilde{L} c$

Proof:

First observe that in virtue of (4) $(a,c) \notin L$ implies $(b,b) \notin L$, which is impossible. The same contradiction derives from $(c,a) \notin L$. So $a L c$ and $c L a$.

LEMMA 2: $\tilde{K} = K \cup \{(a,b);\ a \tilde{L} b\}$ is an equivalence relation A.

Proof:

Reflexivity and symmetry are obvious.
Transitivity: $a \tilde{K} b$ and $b \tilde{K} a$ implies ($a K b$ or $a \tilde{L} b$) and ($b K c$ or $b \tilde{L} c$). Using distributivity of "and" and "or" and conditions (1), (3), (5) and Lemma 1, we get $a \tilde{L} c$ or $a K c$.

LEMMA 3: $a K a$ and $b K b$ iff $a K b$ and $a \tilde{L} b$.

Proof:

Assume $a K a$, $b K b$ and $a L b$; (4) implies $b L a$. Replacing $a L b$ in this assumption by $b L a$, yields $a L b$. So, by the connectivity of L, $a L b$ follows. $a K b$ is now a consequence of (5).
The opposite direction is obvious from (5).

From this Lemma we see that $a K b$ and $a L b$ occurs at best once. So, generally, it is possible to separate the small elements of an equivalence class of K from the large ones. Denoting the equivalence class of a with respect to $\tilde{K}$ by a_K, we define

$$s(a) = \{b \in a_K;\ b L c \text{ for all } c \in a_K\}$$

$$t(a) = \{b \in a_K;\ c L b \text{ for all } c \in a_K\}$$

The properties of $s(a)$ and $t(a)$ are dwelt upon in the next few lemmas.

LEMMA 4: $s(a) \neq \emptyset$. If $b,c \in s(a)$, then $b \tilde{L} c$. The same holds for $t(a)$.

Proof:

Suppose, $b \in a_K$ exists such that $(a,b) \notin L$; then $b K a$. Now $b \in s(a)$, for if there is $b' \in a_K$ $(b,b') \notin L$, then $b' K b$ and $b' K a$, using Lemma 2.

But from Lemma 1 $a \tilde{L} b$ is inferred, contradicting the choice of b.

If there is no such b, then $a \in s(a)$.

Now, if $b,c \in s(a)$, then $b L c$ and $c L b$, proving the second part. A similar proof applies for $t(a)$.

LEMMA 5: $s(a) = t(a) = a_K$ or $s(a) \cap t(a) = \emptyset$.

Proof:

Assume $b \in s(a) \cap t(a)$; then for all $c \in a_K$, $b L c L b$, implying $b \tilde{L} c$. Therefore, $c \in s(a)$ and $c \in t(a)$, which is equivalent to $s(a) = a_K$ and $t(a) = a_K$. If no such b exists, the intersection is empty.

LEMMA 6: $s(a) \cup t(a) = a_K$

Proof:

By Lemma 5 there are only two cases possible, one is trivial. Now, let $s(a) \cap t(a) = \emptyset$ and $b \in a_K$.

We assume $b \notin s(a)$ and show $b \in t(a)$. Since $s(a) \neq \emptyset$, there is $c \in s(a)$; furthermore $(b,c) \notin L$, for otherwise $b \in s(a)$; so $b K c$.

For arbitrary $d \in a_K$, if $d \notin s(a)$, then $d K c$. By Lemma 1, $d \tilde{L} b$.

If $d \in s(a)$, then $d L b$. In any case we have $d L b$, showing $b \in t(a)$, from which the proposition follows.

LEMMA 7: If $b \in s(a)$, $c \in t(a)$, then $b K c$.

Proof:

Since $b,c \in a_K$, $b K c$ or $b \tilde{L} c$. But $b \tilde{L} c$ would imply $c \in s(a)$ and by Lemma 5 $t(a) = s(a) = a_K$. But by (2) $b' \in A$ exists such that $b K b'$, so $b' \in a_K = s(a)$, whence by Lemma 4 $b' \tilde{L} c$ and by (5) $b K c$.

LEMMA 8: $a K b$ iff ($a \in s(a)$ and $b \in t(a)$) or ($a \in t(a)$ and $b \in s(a)$).

Proof:

Assume $a K b$. Of course, $a,b \in a_K$; if $a,b \in s(a)$, then $a \tilde{L} b$ and $s(a) = t(a)$, i.e. $b \in t(a)$; but if $a \in s(a)$ and $b \notin s(a)$, then by Lemma 6, $b \in t(a)$.

A similar argument applies, if we assume $a \in t(a)$. The opposite implication is contained in Lemma 7.

LEMMA 9: If $s(a) = t(a)$ and $s(b) = t(b)$, then $a \tilde{L} b$ and $a K b$.

Proof:

From Lemma 7 we infer $a K a$ and $b K b$ and Lemma 3 then gives the assertion.

Our next step is to define a relation on the set of equivalence classes of $\tilde{K}$, which turns out to be a simple order, i.e. a weak order with indifference implying indentity.

$$A_K = \{a_K;\ a \in A\}$$

$a_K L_s b_K$ iff there are $c \in s(a)$, $d \in s(b)$ such that $c L d$

$a_K L_t b_K$ iff there are $c \in t(a)$, $d \in t(b)$ such that $c L d$.

LEMMA 10: If $a \in s(x)$ and $b \in t(y)$, then $a L b$.

Proof:

Suppose $b L a$ and $a' K a$, $b' K b$. Condition (4) yields $a' L b'$. From $a' K a$ and $a \tilde{K} x$ we infer $a' \tilde{K} x$ and $a L a'$. Similarly $b' L b$. Now $a L a'$, $a' L b'$, and $b' L b$ gives $a L b$. If not $b L a$, then of course $a L b$ holds.

LEMMA 11: a) $s(a) \cap s(b) = \emptyset$ or $s(a) = s(b)$

b) $t(a) \cap t(b) = \emptyset$ or $t(a) = t(b)$

c) $x \in s(a) \cap t(b)$ implies $s(a) = t(a) = s(b) = t(b)$.

Proof:

a) Suppose $x \in s(a) \cap s(b)$. Then $a \tilde{K} b$ and $a_K = b_K$. If $y \in s(a)$, then $y \in b_K$. Moreover $x \tilde{L} y$. Take $z \in b_K$; by definition of $s(b)$ $x L z$, so $y L z$ and $y \in s(b)$.

b) The proof is similar to a).

c) Obviously $x' \in s(b) \cap t(a)$, where $x' K x$. But $x \in s(a)$ yields $x' \in t(a)$, so $x L x'$, whereas $x' \in s(b)$ gives $s' L x$; together $x \tilde{L} x'$. By Lemma 3, $x K x$, by Lemma 5, $s(a) = t(a)$ and $s(b) = t(b)$, and by Lemma 9, $a \tilde{L} b$. Thus $s(a) = s(b)$ and $t(a) = t(b)$.

LEMMA 12: (A_K, L_s) and (A_K, L_t) are simple orders.

Proof:

Since the elements of $s(a)$ are all L-equivalent, the definition of L_s does not depend on the particular elements c, d chosen from $s(a)$, $s(b)$ respectively. Connectivity, antisymmetry, and transitivity are readily shown from the definition.

LEMMA 13: If D is an order dense subset of (A, L), then $D_K = \{d_K;\ d \in D\}$ is an order dense subset of A_K.

Proof:

If $a_K\ L_s\ b_K$, $c \in s(a)$ and $e \in s(b)$ exist satisfying $c\ L\ e$. If, furthermore, $(b_K, a_K) \notin L_s$, then $(e, c) \notin L$. Since D is order dense in A, $d \in D$ exists such that $c\ L\ d$ and $d\ L\ e$.

Now, if $d \in s(d)$, we immediately have $a_K\ L_s\ d_K$ and $d_K\ L_s\ b_K$. But, if $d \in t(d)$, by Lemma 10 $e\ L\ d$, since $e \in s(b)$; hence $d\ L\ e$ and $d \in s(b)$. From Lemma 11c) we have $s(d) = t(d)$, so again $d \in s(d)$.

Now, the proof of Theorem 1 is easily provided. Since D_K specified in Lemma 13 is countable provided D is countable, we have a representation $g\colon A_K \to \mathbb{R}$, monotone with L_s. Then, defining for $a \in A$

$$f(a) = \begin{cases} -\exp(-g(a_K)) & \text{if } a \notin t(a) \\ \exp(-g(a_K)) & \text{if } a \notin s(a) \\ 0 & \text{if } a \in s(a) \cap t(a), \end{cases}$$

where $\exp(x) = e^x$, we eventually obtain the desired representation. By Lemmas 5 and 6 we see that f is well defined for all $a \in A$. The reader may verify (4.1). Concerning (4.2) we observe referring to Lemma 8:
$f(a) + f(a') = 0$ iff $a\ K\ a'$.
The "if" part of this statement brings about (4.2), whereas the "only if" part is reformulated in the following Corollary of Theorem 1.

Corollary. Let (A, L, R) satisfy the assumptions of Theorem 1. Then there is a real number k and a representation f such that (4.1), (4.2) and

$$\text{if } f(a) + f(b) = k, \text{ then } a\ K\ b \tag{4.4}$$

hold.
The constant k is zero in the representation (4.3).

The next theorem specifies the conditions of the factorization and additive decomposition.

THEOREM 2

The system $(A, \leq, L, K)$ consisting of a nonempty set A and three binary relations $\leq$, L, K on A satisfying

(1) $(A, \leq)$ is a weak order

(2) (A, L) is a quasiorder

(3) If $a \sim b$, then $a\ L\ b$ or $b\ L\ a$

(4) If $a\ K\ b$, then $a \sim b$

(5) $(\bar{a}, L_{/\bar{a}}, K_{/\bar{a}})$ is a compensatory structure

(6a) For all $a,b \in A$ there is at least one element c such that $a \sim c$ and $b\ \tilde{L}\ c$

(6b) For all $a,b \in A$ there is at least one element c such that $a\ \tilde{L}\ c$ and $b'\ \tilde{L}\ c'$, where $b\ K\ b'$ and $c\ K\ c'$

(6c) There are $a,b \in A$ such that $a \leq b$

(7) If $a\ L\ b$, $a\ K\ a'$, $b\ K\ b'$, $a'\ L\ b'$, then $a \leq b$.

Then, there are sets X_1, X_2 and surjective maps $\eta_i: A \rightarrow X_i$ $(i=1,2)$ and a weak order $\leq'$ on $X_1 \times X_2$ such that

$$a \leq b \text{ iff } \eta_1(a)\ \eta_2(a) \leq' \eta_1(b)\ \eta_2(b). \tag{4.5}$$

Moreover, if a suitable Archimedean condition holds, then $(X_1, X_2, \leq')$ is an additive conjoint structure.

In the latter case there are two functions $f_i: A \rightarrow \mathbb{R}$; $i=1,2$ such that

$$a \leq b \text{ iff } f_1(a) + f_2(a) \leq f_1(b) + f_2(b) \tag{4.6}$$

If $a\ L\ b$ then $f_1(a) \leq f_1(b)$

$$a\ K\ b \text{ iff } f_1(a) = f_2(b) \text{ and } f_2(a) = f_1(b). \tag{4.7}$$

A few comments on the conditions are in order.

First, we want to bring the few properties displayed in chapter 3 and the conditions of both theorems into relation. An indifference class of a conjoint structure is a compensatory structure. Especially (3.2) and condition (4) coincide. An equivalent to L' of a conjoint structure can easily be defined in $(A, \leq, L, K)$ availing oneself of (3.3), i.e. $a\ L'\ b$ iff $a\ K\ a'$, $b\ K\ b'$ and $a'\ L\ b'$. Condition (7) of Theorem 2 seems important as it takes care of the compatibility between $\leq$, L, and K. It originates from (3.4).

Second, we do not insinuate the necessity of our conditions with this comparison. This problem is not yet dealt with properly. For example the existential axioms (6) are far from being necessary. On the contrary they force the structure to be so bulgy that the representation cannot result even in a bounded rectangle, let alone an incomplete structure, which are treated by Doignon, Falmagne (1974), Jaffray (1974), Narens, Luce (1976), Suck (1983) and which are very desirable from the point of view of applications.

Third, an extension to more than two components would be expedient. Provisionally, one can imagine the indifference classes of $\leq$ to consist of a structure that satisfies Theorem 2. Somehow, an iterative application of Theorem 2 should produce a multidimensional structure, but as yet we cannot present a final answer.

Fourth, because of the deficiencies just mentioned and the amount of technical lemmas needed, the proof of Theorem 2 is omitted.

5. REMARKS ON APPLICATIONS

The difficulties that arise in applying our theorems to empirical work in psychology must not be shoved under the rug. The data have to provide information about an order relation with indifference sets that are sufficiently large to disclose a finer structure, which is a candidate for the L and K relations. One should at least have a good guess about one of the components. In this case the experiment can possibly be planned sophisticatedly enough to produce the necessary information.

Although the possibilities of applications are rather poor, there are several problems of theoretical interest, which can be reformulated in terms of the present development.

One example is the question of the foundations of multidimensional scaling. Cast in the framework of the present paper it turns out somewhat like this: given a set of stimuli A and a proximity relation P; assume P is a weak order on $A \times A$. Do sets $X_1, \ldots, X_n$ and a weak order P' exist such that $(A \times A, P)$ and $(X_1 \times \ldots \times X_n, P')$ are homomorphic and is there a representation, which is compatible with a metric? An analogue of Theorem 2 for n components could specify sufficient conditions. The approach of Beals, Krantz, Tversky (1968) (see also Beals, Krantz (1967) and Tversky, Krantz (1970)) sidesteps the question of the existence of the factors, because it starts with the factorial structure $(X_1, X_2, \ldots, X_n, P')$. The papers of Drösler (1979) using axioms from abstract geometry to characterize a plane and Orth (1980) determining one-dimensionality in $(A \times A, P)$ are more in tune with a component free axiomatization. The foundational problem of multidimensional scaling has raised a partially controversial debate (see for example Roskam (1981) and Schönemann, Borg (1981)), and a satisfactory solution is not yet achieved.

Another line of the investigation of similarity data focused feature matching. Tversky (1977) developed a feature theory based on the assumption that each object of the domain under study can be replaced by a set of features taken from a sufficiently large reservoir. Denoting the feature set associated with the object a by A and the similarity index by s (a,b), it is shown that a similarity scale S and a nonnegative function f exists satisfying

$$\begin{aligned} s(a,b) \leq s(c,d) &\text{ iff } S(a,b) \leq S(c,d) \\ S(a,b) = \theta f(A\cap B) - \alpha f(A-B) - \beta f(B-A) &\text{ for some } \theta,\ \alpha,\ \beta \geq 0 \end{aligned} \qquad (5.1)$$

This representation does considerably better in case of asymmetric similarities and other pecularities that cannot easily be reconciled with the properties of a metric space.

The representation (5.1) is derived from a complicated set of axioms including the existence of the feature set and the following matching condition

$$s(a,b) = F(A\cap B,\ A-B,\ B-C) \qquad (5.2)$$

where F(X, Y, Z) is a function of three arguments, each a subset of the feature set. Although intuitively appealing, this seems to be a strong condition suggesting in a sense the decomposability of the similarity into the relevant factors, which are the common features $A\cap B$, the features a has and b not, A-B, and b has and a not, B-A.

We conjecture that our formulation of the problem of multidimensional scaling just explained and an application of a proper variant of Theorem 2 could yield a representation (5.1) without using (5.2).

Further applications are conceivable in the investigation of the structure of interaction effects in factorial designs (Krantz et al. (1971), p.248 and Fishburn (1975)).
But in this as well as in the aforementioned problems future research has to show, whether our expectations of fruitful applications will bear out.

REFERENCES

[1] Beals, R., Krantz, D.H., Metrics and geodesics induced by order relations, Mathematisches Zeitschrift, 101 (1967), 285-298.

[2] Beals, R., Krantz, D.H., Tversky, A., Foundations of multidimensional scaling, Psychological Review, 75 (1968), 127-142.

[3] Doignon, J.-P., Falmagne, J.C., Difference measurement and simple scalability with restricted solvability, Journal of Mathematical Psychology 11 (1974), 473-499.

[4] Drösler, J., Foundations of multi-dimensional metric scaling in Cayley-Klein geometries, British Journal of Mathematical and Statistical Psychology, 32 (1979), 185-211.

[5] Fishburn, P.C., Nondecomposable conjoint measurement for bisymmetric structures, Journal of Mathematical Psychology, 12 (1975), 75-89.

[6] Holman, E.W., A note on conjoint measurement with restricted solvability, Journal of Mathematical Psychology, 8 (1971), 489-494.

[7] Jaffray, J.Y., On the extension of additive utilities to infinite sets, Journal of Mathematical Psychology, 11 (1974), 431-452.

[8] Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A., Foundations of measurement, Vol. 1, New York, Academic Press (1971).

[9] Narens, L., Luce, R.D., The algebra of measurement, Journal of Pure and Applied Algebra, 8 (1976), 197-233.

[10] Orth, B., On the foundations of multidimensional scaling: An alternative to the Beals, Krantz, Tversky apporach, In Lantermann, E.D., Feger, H. (Eds.), Similarity and Choice, Papers in honour of Clyde Coombs, Bern, Huber (1980).

[11] Roberts, F.S., Measurement theory with applications to decisionmaking, utility and the social sciences, Reading, Mass., Addison-Wesley (1979).

[12] Roskam, E.E., Contributions of multidimensional scaling in social science research, In Borg, I. (Ed.), Multidimensional Data Representation: when and why, Ann Arbor, Mathesis Press (1981).

[13] Schönemann, P.H., Borg, I., Measurement, scaling and factor analysis, In Borg, I., Multidimensional Data Representation: when and why, Ann Arbor: Mathesis Press (1981).

[14] Suck, R., On the interrelation between extensive measurement and additive conjoint measurement, In Micko, C., Schulz, U., Formalization of psychological theories. Proceedings of the 13th Eur. Math. Psychol. Group Meeting, Bielefeld: Materialien des Universitätsschwerpunktes Mathematisierung XXXX, (1983).

[15] Tversky, A., Features of similarity, Psychological Review, 84 (1977), 327-352.

[16] Tversky, A., Krantz, D.H., The dimensional representation and the metric structure of similarity data, Journal of Mathematical Psychology, 7 (1970), 572-596.

PART III

DATA ANALYSIS

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

THE PROBLEM OF REPRESENTATION BASED UPON TWO CRITERIA

G. De Meur, M. Gassner, X. Hubaut

Université Libre de Bruxelles

The problem considered here is that of electing a delegation when the population of electors is separated into categories, the division being based upon one or several criteria.
A two criteria-based situation currently exists in the Netherlands. Using a slightly more general version of this system, we proved that the only type of situation avoiding impossibilities is one where each criterion only presents two alternatives. Furthermore, setting a number of conditions, as weak as possible, which must be respected in order to ensure fair representation, the same result remains - which is not so surprising as it could have been suggested by Arrow's famous theorem.

1. INTRODUCTION

A population must elect a certain number of delegates who will form an assembly such as the Parliament or the Congress,...

The problem of this representation being fair or not arises because the population is divided into categories, the division being based upon one or several criteria. These criteria are usually geographic (provinces in Belgium, départements in France, states in the U.S.,...) and partisan (when the population is divided into political categories), but other types of motives can occur (and have along the years) such as economic, social or even racial criteria.

The perfect and ideal solution would be to elect a delegation proportionally (in the mathematical sense) to the number of persons in each of the different categories; but since integer solutions are, of course, required, the problem is practically always impossible. It is thus necessary to find a distribution of seats with as little variation as possible from the different units' quotas ("number of delegates deserved by the strict proportional rule) and which does not transgress our natural sense of fairness.

For this purpose, one must first establish a number of rules, which, when being followed, should ensure fair representation. The second step is to find apportionment methods which obey the given imperatives. The problem, when it involves one criterion, has been thoroughly discussed, and a general review of the research made in that field can be found in Balinski and Young [2].

To our best knowledge, except for a paper by Anthonisse [1], examining two-criteria cases, the same kind of problem involving more than one criterion has never been considered though these cases are frequent in many countries. We focussed our attention on this type of problem expanding it to a somewhat more general frame.

2. THE ONE-CRITERION PROBLEM

Generalizing the one-criterion problem, part of its structure remains; it will thus be interesting and useful to review some of its specific aspects. What can we expect to happen - or rather what can we reasonably hope to avoid - in a one-criterion based situation?
Several "different" paradoxes can occur and thus brings to the mind some "natural" questions (and derived rules) such as:
(i) Can we accept an apportionment method which allows a party to receive more seats than its quota rounded up?
(ii) For a determined number of delegates, if one class of the population increases while another decreases, can we let the first's apportionment decrease while the second's increases?
(iii)Should a method "systematically favor" small units at the expense of large ones or vice-versa?
(iv) Can we allow the creation of a new unit to change the relative positions of the others?
(v) If the total representation size increases, can a unit lose any delegates (if, of course, no outside reason for this should appear)?

Answering the first question, some systems do accept such methods for a very logical reason; as a matter of fact, one of them is currently used in the Belgian "legislative" elections [3].

Simplifying slightly, this is how it works.

Suppose a population of 100000 must elect 100 representatives. The vote is cast and parties A, B and C receive respectively 2920, 95220 and 1860 votes. The first stage is to establish each party's quota and to give out "direct" seats (whole numbers contained in quotas) (fig. 1a).

This way, 98 seats have been assigned. Which party (ies) should get the two remaining seats? In a way to the party(ies) "offering the most". By this, we mean the following: fictively, one extra delegate is added to each party's total and, for each party, we compute the "value" of a delegate (number of voters per delegate). The party whose representatives are worth the most voters receives the first extra seat; here, it is party B (fig. 1b).

One more seat has not yet been assigned. The same system is used and it is again B who receives it, and thus finds itself with 97 seats, the surprising prediction (fig. 1c).

PARTY	VOTES	QUOTA	DIRECT SEATS
A	2920	2.92	2
B	95220	95.22	95
C	1860	1.86	1
	100000	100	98

Figure 1a

PARTY	DIRECT SEATS + 1	"VALUE" OF EACH DELEGATE	SITUATION 1
A	3	973.3	2
B	96	$\boxed{991.8}$	96
C	2	930	1
			99

Figure 1b

PARTY	(SITUATION 1) + 1	"VALUE" OF EACH DELEGATE	FINAL REPRESENTATION
A	3	973.3	2
B	97	$\boxed{981.6}$	97
C	2	930	1
			100

Figure 1c

Using one of the first methods used in the United States (named after its creator, Alexander Hamilton), the type of situation described in point (ii) can occur. Hamilton's method consists in rounding the units' quotas, giving to each unit the whole number contained in its quota and then assigning any remaining seats to those having the largest remainders.

A simple example of the situation can be found in figure 2.

17 SEATS AVAILABLE

t

POPULATION	QUOTA	DELEGATES
275	4.68	5
177	3.01	3
393	6.68	7
155	2.64	2
1000		17

→

HAMILTON'S METHOD

t'

POPULATION	QUOTA	DELEGATES
304	4.95	5
190	3.09	3
395 (↑)	6.44	6 (↓)
154 (↓)	2.51	3 (↑)
1043		17

Figure 2

3. THE DUTCH PROBLEM

What can we hope to happen if the problem involves more than one population-division criterion? A two-criteria based situation currently existing in the Netherlands is the motivation of this study.

Voters of a region are characterized by the commune in which they live and by the party of their choice.

The first step of the election is for the population to vote for commune councillors.

The second step is for all councillors of a region to form a regional

council composed of a number of their members.
The Dutch law states:
"Each communal council assigns one quarter of its members to the regional council, fractions of .5 or more being rounded up, fractions under .5 being *rounded down*".

The legislator also forecasts that an effort should be made so the political composition of the regional council should match that of the region's communal councils.

The situation can be mathematized as follows:

consider a matrix S where element s_{ij} is the number of party P_j's councillors in commune C_i. The regional council is described by a matrix R, "derived" from S in the sense that element r_{ij} is always smaller or equal to the corresponding element s_{ij}. This matrix R must also respect imperatives of political and regional balance. To be more precise:

1) The sum $r_{i.}$ of the elements of line i must coincide with the number of regional delegates assigned to commune C_i, as established by the legal rule mentioned above.
2) The sum $r_{.j}$ of the elements of column j must be equal to the number of representatives assigned to party P_j as determined by Hamilton's method (fig. 3a and 3b).

How can one establish this matrix R?

The first idea entering in mind would be to treat each commune separately, as if it was the only one, assigning the representatives by Hamilton's method. One calls this the initial situation (fig. 3c). By means of construction, each commune receives its fair share of representatives. Nevertheless this omits to take global "party balance" into consideration because nothing in the construction guarantees that the total number of delegates-per-party should be exactly the one established beforehand by Hamilton's method. Yet, it is possible that these numbers coincide. In that case, the initial matrix is the matrix R. In general, however, these do not coincide. How can the initial matrix be altered in order to obtain one that satisfies the two conditions simultaneously?
Because the total number of representatives by commune is predeterminated, transferring seats can only be authorized between parties within the same commune. Obviously it is better to stay as close as possible to the initial

setting.

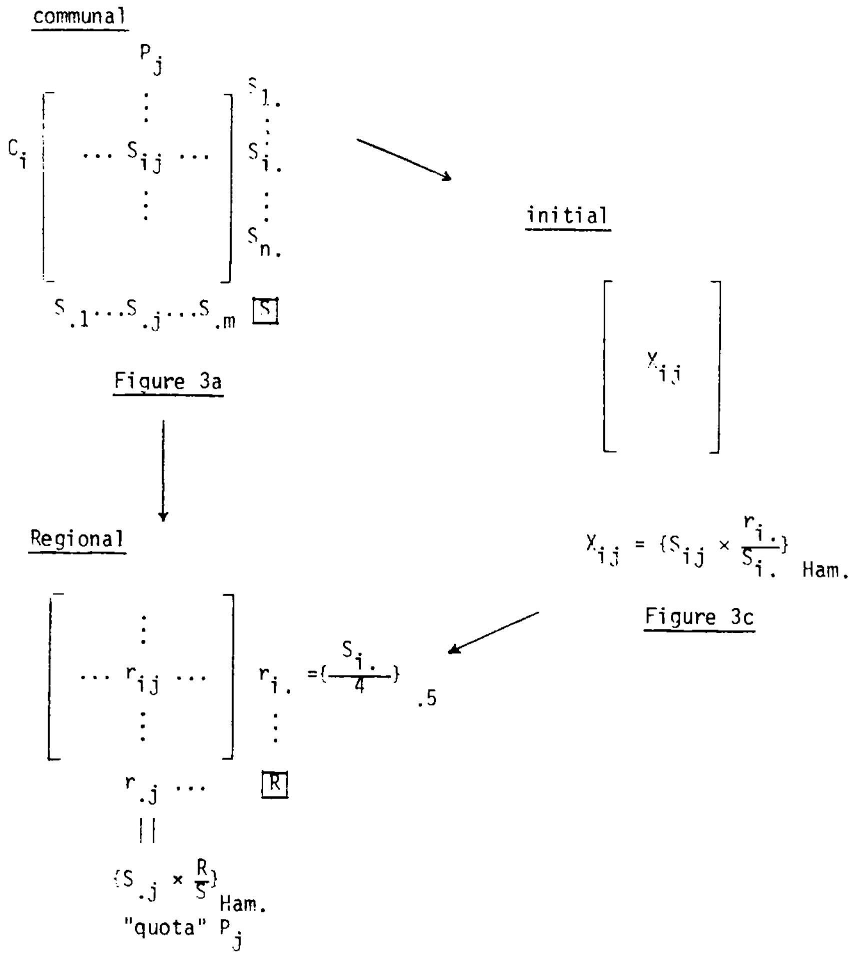

communal
P_j
C_i
$\dots S_{ij} \dots$
$S_{1.}$
$S_{i.}$
$S_{n.}$
$S_{.1} \dots S_{.j} \dots S_{.m}$
$\boxed{S}$
initial
X_{ij}
$X_{ij} = \{S_{ij} \times \frac{r_{i.}}{S_{i.}}\}_{\text{Ham.}}$
Regional
$\dots r_{ij} \dots$
$r_{i.} = \{\frac{S_{i.}}{4}\}_{.5}$
$r_{.j} \dots$
$\boxed{R}$
$\{S_{.j} \times \frac{R}{S}\}_{\text{Ham.}}$
"quota" P_j

Figure 3a

Figure 3c

Figure 3b

In certain cases, the alterations can be performed without too much ambiguity as shown by the example of figure 4.

communal

	P_1	P_2	P_3	P_4	$\Sigma_\downarrow$	
C_1	5	2	3	9	19	5
C_2	0	1	1	5	7	2
C_3	7	2	0	3	12	3
C_4	4	0	0	3	7	2
C_5	0	3	7	3	13	3
$\Sigma_\rightarrow$	16	8	11	23	58	15

initial

1	1	1	2	5
0	0	0	2	2
2	0	0	1	3
1	0	0	1	2
0	1	1	1	3
4	2	2	7	15

regional

1	1	1	2	5
0	0	0	2	2
2	0	0	1	3
1	0	0	1	2
0	1	$\boxed{2}$	$\boxed{0}$	3
4	2	$\underline{3}$	$\underline{6}$	15

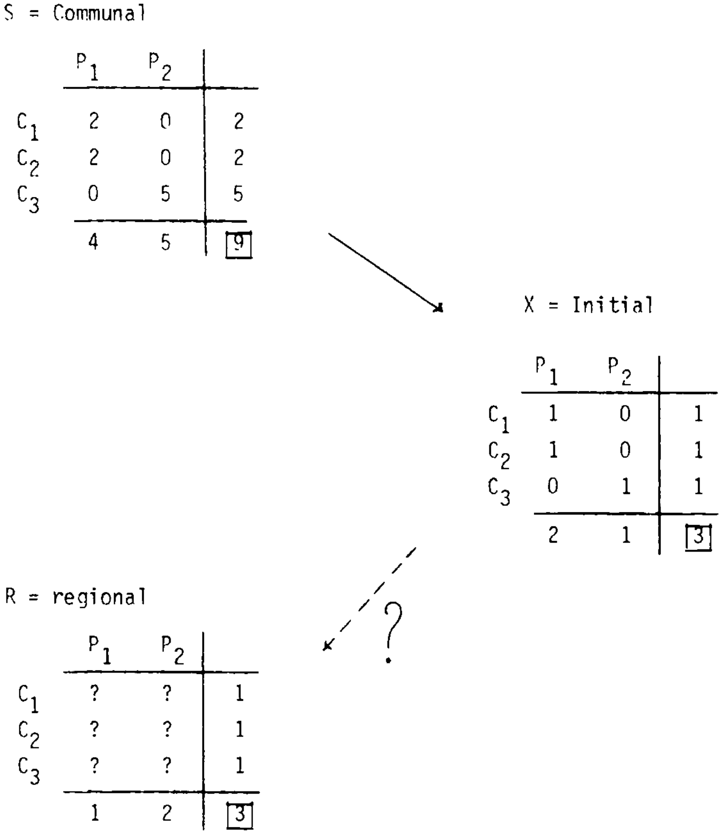
Figure 4

The apportionment method seems to be well established now. There are, however, some unsolvable cases, as can be seen in fig. 5.

The initial setting gives two seats to party P_1 and only one to party P_2. Global political balance, however, imposes two to P_2 and one to P_1. P_1 should give up one seat to party P_2, which is clearly impossible.

We generalized the problem using divisor N instead of 4 and proved that the only situation avoiding impossibilities is one where there are only two parties and two communes.

We tried to see if the impossibilities in the other cases were due to the particularity of the system used in Holland. They appeared to be more fundamental.



S = Communal

	P_1	P_2	
C_1	2	0	2
C_2	2	0	2
C_3	0	5	5
	4	5	$\boxed{9}$

X = Initial

	P_1	P_2	
C_1	1	0	1
C_2	1	0	1
C_3	0	1	1
	2	1	$\boxed{3}$

R = regional

	P_1	P_2	
C_1	?	?	1
C_2	?	?	1
C_3	?	?	1
	1	2	$\boxed{3}$

Figure 5

4. CRITICISM OF THE SYSTEM

One could think that the impossibility discovered here is the direct result of the method used in rounding the political quotas (Hamilton's method). Indeed, why choose to use a method which has been proven to raise various paradoxes?

Furthermore, by accepting transfers of seats from party to party in the same commune, the distribution resulting from the method used in the first place can be destroyed. Going back to the building of the matrix R, we asked ourselves:

a) which of its characteristics are predetermined?

b) which requirements (as weak as possible) should be preimposed upon the elements of R?

Answering the first question, it is obvious

1) that the sum of the elements of each line can be considered predetermined; furthermore, these sums are not determined arbitrarily: the order created by the number of communal councillors of each commune must be respected. Also, the difference between the numbers of regional delegates cannot exceed the difference between the numbers of their communal councillors (except in case of a tie).

Symbolically:

$$\forall i,j \quad s_{i.} < s_{j.} \Rightarrow r_{i.} \leq r_{j.} \text{ and } s_{j.}-s_{i.} \geq r_{j.}-r_{i.} \quad .$$

2) that the sum of the elements of each column of R is predetermined, this time by some apportionment method (not necessarily Hamilton's).

We must have:

$$\forall i,j \quad s_{.i} < s_{.j} \Rightarrow r_{.i} \leq r_{.j} \text{ and } s_{.j}-s_{.i} \geq r_{.j}-r_{.i}$$

in order to respect global political balance.

Concerning the requirements to be imposed upon R's elements, the following condition seems natural:

$$\forall i,j,k,l \quad s_{ij} < s_{kl} \Rightarrow r_{ij} \leq r_{kl}$$

Party P_l in commune C_k, having more councillors than party P_j in commune C_i, should send at least as many delegates as P_j to the regional council. The previous condition, however, seems to be very strong and perhaps artificial in the sense that it is based upon a comparison between absolute numbers while the problem fundamentally demands a relative balance of those numbers based upon both the communes and the parties. Let us replace it by the following weaker and more natural condition

$$\forall i,j,k \quad s_{ij} < s_{ik} \Rightarrow r_{ij} \leq r_{ik} \text{ and } s_{ik}-s_{ij} \geq r_{ik}-r_{ij}$$

requiring that the order created by the number of communal councillors per party within the same commune will not be reversed in the regional council. By not respecting a similar condition on the columns of R, we are less in contradiction with common political morals than by violating the above mentioned condition imposed upon the communes.

Unexpectedly the example, shown in figure 5, does not satisfy these weak demands.
Which conlusions can be drawn?
Possibly this example may be too particular:
1°) it is not 'connected' in the sense that the set of communes can be separated into two groups $\{C_1,C_2\}$ and $\{C_3\}$ in such a way that no party has a seat in both groups at the same time.

2°) the presence of zeroes, especially while dealing with votes seems very improbable.

These rightful objections can be opposed by the following examples (fig. 6).

An easy, but tedious, verification will allow us to show that the establishment of matrix R is always feasible in the elementary case of two communes and two parties, and that these are thus the only feasible cases.

To end, it is finally not so surprising that everything works better when each criterion only presents two alternatives, as could have been suggested by Arrow's famous theorem and also by political reality, the last round of many elections often being held between two opponents.

(3,2) EXAMPLE

S

4	1	5
2	3	5
2	3	5
8	7	15

$\longrightarrow$

R

A	B	1
C	D	1
E	F	1
2	1	3

with $A \geq B$

$C \leq D$

$E \leq F$

$\Downarrow$

$A = 1$ and $B = 0$

$C = 0$ and $D = 1$

$E = 0$ and $F = 1$

which leads to a contradiction since $A+C+E = 1$ and $B+D+F = 2$.

(2,3) EXAMPLE

S

3	4	1	8
3	1	4	8
6	5	5	16

$\longrightarrow$

R

	A	B	C	1
	D	E	F	1
	1	1	0	2
or	1	0	1	
or	2	0	0	

with $C \leq A \leq B$

and $E \leq D \leq F$

$\Rightarrow B = 1$ and $A = C = 0$

$F = 1$ and $D = E = 0$

which leads to a contradiction since $A+D = 0$ and $B+E = C+F = 1$.

Figure 6

REFERENCES

[1] Anthonisse, J.M., Geografische en politieke samenstelling van de Gewestraad, to appear, presented in Brussels (1981).

[2] Balinski, M.L., and Young, H.P., Fair representation meeting the ideal of one man, one vote, Yale University Press, New Haven and London (1982).

[3] Code Electoral, Moniteur du 19/8/1928 et Moniteur du 28/4/1929.

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

TREE REPRESENTATIONS OF RECTANGULAR PROXIMITY MATRICES

Geert De Soete
University of Ghent, Belgium

Wayne S. DeSarbo, George W. Furnas, and J. Douglas Carroll
Bell Laboratories, Murray Hill

A least-squares algorithm for fitting ultrametric and path length or additive trees to two-way, two-mode proximity data is presented. The algorithm utilizes a penalty function to enforce the ultrametric inequality generalized for asymmetric, and generally rectangular (rather than square) proximity matrices in estimating an ultrametric tree. This stage is used in an alternating least-squares fashion with closed-form formulas for estimating path length constants for deriving path length trees.

1. INTRODUCTION

One often encounters two-way, two-mode $n \times m$ asymmetric rectangular matrices $(\Delta = \|\delta_{ij}\|)$ of proximity data in the behavioral (and other) sciences. They indicate the relationships between two different classes of entities (e.g., objects and variables, subjects and stimuli, stimuli and responses). The analysis of such data often seeks understanding of the structure of these sets of objects and their inter-relationships. For example, profile data of a stimulus by scale type are often formed when one evaluates or rates a number of different stimulus objects on various attribute scales, e.g. a number of different soft drinks are rated on various prespecified scales relating to taste, level of carbonation, etc. If these ratings are interpreted as judgments of closeness to an ideal exemplar of the attribute being rated, the resulting matrix can be viewed as a rectangular proximity matrix.

Another typical example of such rectangular data occurs where one collects subject x stimulus data. For instance, various subjects render preference judgments over a number of different stimuli (e.g. subject preference ratings for various brands of telephones).

Still another typical type of nonsymmetric data encountered in the behavioral sciences are confusions data where $n = m$ (so that Δ is square) and the row and column elements correspond to the same objects. Yet, δ_{ij} is not necessarily equal to δ_{ji}. For example, the row elements might be various messages sent and the column elements the various messages received. The main diagonal elements would record the accuracy of the communication process, while the off-diagonal elements would reflect the error of the process. The famous Miller-Nicely (1955) data on confusions of consonants phonemes in a variety of noise conditions provide an excellent example of such proximities. Also, one could consider brand-switching matrices such as those presented in DeSarbo (1982) where probabilities of switching from brand/product i in period t to j in period $t+1$ represent the δ_{ij} entries in Δ. Here the δ_{ii} entries represent brand loyalty and the δ_{ij} $(i \neq j)$ entries reflect the degree of brand switching. Similar types of matrices are found in the sociological literature in the form of social interaction and mobility matrices.

In summary, there are a number of different types of rectangular data found across many diverse disciplines. We propose here a new methodology for fitting ultrametric and additive tree structures to such data. We begin with a review of the related literature in two-way, two-mode clustering. The new proposed model and algorithm are then described. An application of fitting an ultrametric tree to brand switching data for various major brands of soft drinks is presented. Finally, future research opportunities in this area are described.

2. PREVIOUS WORK IN TWO-MODE CLUSTERING

There exist a number of quite different approaches to clustering both row and column elements in a two-way, two-mode rectangular data matrix. Tryon and Bailey (1970) describe a heuristic for examining both variables and individuals in a clustering setting. Here, a "V-analysis", or "clustering" of the variables (columns), is performed initially via a factor analysis on the variable by variable correlation matrix or on a subset of the most collinear variables. Then, an "O-analysis" of the individuals is performed where subjects are initially scored on the several clusters/factors obtained in the "V-analysis" and a clustering of the subjects is then performed.

One finally obtains a reduced "space" for variables and one for subjects based on their scores from the "V-analysis". The user can then examine profile scores of the various clusters. Thus, the Tryon and Bailey (1970) procedure does not really provide a *joint* clustering of subjects and variables, but rather a clustering of subjects whose resultant clusters vary continuously on a set of factor scores.

Hartigan (1975, 1976) and Hartigan and Engelman (1983) develop a block clustering method for categorical rectangular data. Each block is defined by a cluster of cases and a cluster of variables such that each variable in the block is constant over the cases in the block, except for cases that also belong to other blocks (blocks may overlap). The constant value taken by a variable in a block is called the "modal value" for that block. The goal of the analysis is to provide a succinct representation of the data by a few large blocks with corresponding block modal value, together with residual single blocks consisting of single values deviating from the appropriate block modal value. An iterative algorithm in Hartigan (1975, 1976) is utilized to cluster both objects and variables simultaneously.

McCormick, Schweitzer, and White (1972) propose their "bond energy algorithm" to provide a joint clustering of row and column elements in rectangular data. This is accomplished by permuting the rows and columns of an input data array in such a way as to push the numerically larger array elements together (or as the authors mention, to maximize the summed bond energy over all row and column permutations in the input array - the "beta measure of effectiveness"). An iterative sequential-selection heurisitic is utilized to provide an at least locally optimum solution.

DeSarbo (1982) has developed the GENNCLUS methodology for the simultaneous clustering of both row and column elements. He generalizes the ADCLUS (Shepard and Arabie (1979)) model, representing interstimulus proximities as combinations of discrete and possibly overlapping properties, to the case of asymmetric proximities. The GENNCLUS procedure allows for the estimation of either overlapping or nonoverlapping clusters. It utilizes a series of gradient based procedures and combinatorial optimization methods in an alternating least-squares framework. Sarle (Note 6) has recently developed a similar procedure for the GENNCLUS "dual-domain" case where separate clusters would be derived for both row and column objects, and be subsequently related to each other via an estimated weights matrix.

None of the above mentioned methods for jointly clustering row and column elements in rectangular data have dealt with fitting tree structures (ultrametric and/or additive trees) to such data. While such methods (cf. Hartigan (1967); Carroll and Chang (1973); Cunningham (1974); Sattath and Tversky (1977); Carroll and Pruzansky (1975); Carroll (1976); Carroll and Pruzansky (1980); De Soete, Note 2) have been developed for fitting such structures to the one-mode symmetric proximities (see Carroll, Clark, and DeSarbo, Note 1, for the three-way case), only Furnas (Note 3) has investigated the "tree-unfolding" problem of fitting a tree to rectangular proximities, providing a joint representation of both row and column objects. He develops an "ultrametric inequality" condition for such rectangular data (he also considers additive trees) and proposes an agglomerative clustering method to fit an ultrametric tree.

We shall extend the work of Furnas (Note 3) to accommodate both ultrametric and additive trees by utilization of a mathematical programming approach based on a penalty function algorithm. The approach to be described here is aimed at explicitly optimizing a least-squares loss function, whereas Furnas' (Note 3) earlier approach was heuristic in nature and only approximately least-squares.

3. THE PENALTY FUNCTION APPROACH

A. THEORY UNDERLYING THE "TREE UNFOLDING" MODEL

The underlying model and theory for unfolding via ultrametric and path length or additive trees was worked out by Furnas (Note 3). He derived necessary and sufficient conditions, and uniqueness properties, for such representations and devised some data analytic methods. Those methods were only heuristics (agglomerative algorithms for Ultrametric trees; ad hoc methods for error-free Additive tree data). The purpose of the current work is to devise and evaluate explicit least squares methods for fitting these models, and give examples of their usefulness.

In its simplest form, usual proximity analysis begins with a single set of objects and a square symmetric matrix of all the pairwise measurements between them. The goal is to find a single global structure (e.g., a spatial configuration or tree structure) representing all the objects and, as faithfully as possible, the proximities between them.

Unfolding analysis is a slight variant. It begins with *two* separate classes of objects (e.g., people and cars) and with the rectangular matrix of measurements only between pairs of opposite classes (e.g, how much each person likes each car). Its goal is still the same - to find a single global representation, placing in it both sorts of objects such that the distances between them represent, as faithfully as possible the original between-class data. As in the usual one-class analysis, the resulting structure represents the data with fewer, hopefully interpretable parameters. An useful by-product of an unfolding analysis is that the global structure provides a representation for the implicit within-class proximities, in addition to the between-class relationships it explicitly tried to fit. Thus for example it becomes evident which cars are similar, by virtue of being liked by the same people, and which people are similar by virtue of liking the same cars.

Unfolding theory has been well elaborated for multidimensional models, but there is good evidence that the similarity structure of many familiar domains is better fit by tree structures (Sattath & Tversky (1977); Carroll, Pruzansky & Tversky (1982)). One would expect unfolding analyses in such domains to require tree representation as well (see for example Furnas, Note 3).

As background to the work presented here, we review the necessary and sufficient conditions for, and the resulting uniqueness of the two-mode tree representations. A brief summary is presented here (for the errorless case) - *further details may be found in Furnas (Note 3).*

Ultrametric distances (i.e., distances associated with ultrametric trees) must obey the Ultrametric Inequality:

$$d_{ij} \leq \max\,(d_{ik}, d_{jk}). \qquad (1)$$

In the case of rectangular distance matrices, however, it is not possible to test satisfaction of the ultrametric inequality since one of the three distances will be missing for every triple. Furnas showed that, for rectangular matrices with a distance measure defined only between items of different classes (A, with elements represented by letters early in the alphabet and Z represented by letters late in the alphabet) the following Two-Class Ultrametric condition is necessary and sufficient for representation as an ultrametric tree:

$$t_{az} \le \max\,(t_{ax}, t_{bx}, t_{bz}) \,. \tag{2}$$

When this inequality is satisfied, the representation is unique up to the internal structure of purely one-class subtrees. For example, the three trees in Figure 1 are equivalent in their two-class structure.

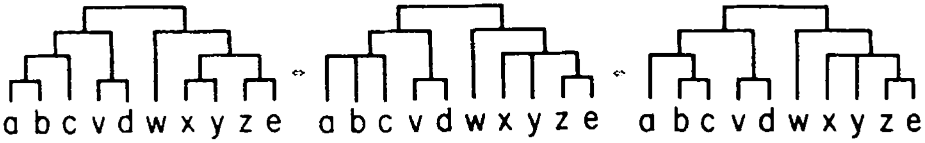


Figure 1. Indeterminacies in Two-Class Trees: Internal Structure of One-Class Clusters.

For example, the (vd) and (ze) subtrees are well determined, but the (abc) and (xy) subtrees are not. By default, such two-class trees will be shown in their least structured form - that is with no structure for the pure subtrees and the fewest possible number of extra nodes, as in the middle example above. Note that this indeterminacy is usually small, in that most subtrees have mixed membership, particularly once they get to be of reasonable size. In the example above, only four of the 45 distances in the final tree were not completely determined, and even those four were strictly bounded.

Necessary and sufficient conditions for unfolding path length or additive trees are much more complicated (Furnas, Note 3, presented a bounded deterministic algorithm for this) and will not be given here. It suffices instead to note that overall distances in a path length or additive tree can be decomposed into the sum of an ultrametric part and an additively decomposable part (see Carroll and Pruzansky (1980)). The same is true for rectangular submatrices of distances from a path length or additive tree, and the analytic techniques proposed here make use of this decomposition.

In path length or additive trees, the unfolded representations have two sorts of non-uniqueness. One is exactly equivalent to the ultrametric case - purely one-class subtrees have indeterminable internal structure. The second is a single continuous free parameter, shifting relative lengths of the terminal branches of the two classes. By adding a constant to all terminal branches of one class and subtracting it from the terminal branches of the other class, all between-class distances have a net change of zero. Thus, for example, the three trees displayed in Figure 2 are compatible with the same two-class matrix.

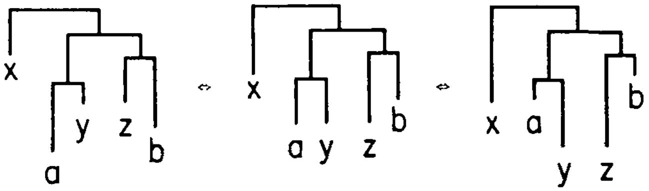


Figure 2. Indeterminacy in Two-Class Additive Trees: Adding a Constant to Terminal Branches of one class and subtracting it from the other.

A convention is used in this paper that attempts to balance the length of the terminal branches of the two classes in a manner to be described later.

These indeterminacies require a degree of caution so as not to "over-interpret" the tree representations resulting from an unfolding analysis, just as the orientation of axes cannot be determined in a two-way multidimensional scaling analysis.

One important point to note is that the approach discussed in this paper makes a very strong assumption about the data (Δ) -- namely that of comparability of the data from row to row (and/or column to column) of the data matrix. In the terminology coined by Coombs (1964), the matrix is regarded as an *unconditional* (rather than a row or column conditional) proximity matrix.

B. THE ALGORITHM

1. For estimating an Ultrametric Tree

The algorithm consists of the following phases.

a) Transform the data matrix Δ into a matrix T best approximating Δ in a least-squares sense where T satisfies the two class ultrametric inequality:

$$t_{i\ell} \leq \max\ (t_{ik}, t_{jk}, t_{j\ell}), \qquad (3)$$

for $i,j = 1,\ldots,n$ $(i \neq j)$ and $k,\ell = 1,\ldots,m$ $(k \neq \ell)$. An alternative statement of the two-class ultrametric inequality which can be shown to be equivalent to expression (3) is that for every quadruple of points comprised of two from each class, the two largest of the four defined distances must be equal, i.e. given $t_{i\ell}, t_{ik}, t_{j\ell}$, and t_{jk} (the only four distances among i and j in class one and k and ℓ in class two defined by the rectangular proximity data), the two largest of those four distances must be equal. The problem can be reformulated as that of solving the optimization problem:

$$\min \{L(T) = \sum_{i=1}^{n} \sum_{k=1}^{m} (\delta_{ik} - t_{ik})^2\}, \tag{4}$$

subject to the condition that T satisfies the two-class ultrametric inequality. To do this, an exterior penalty function approach (Rao (1972)) is utilized to convert the constrained problem into a series of unconstrained ones. The augmented function:

$$\Phi(T,\rho) = L(T) + \rho P(T), \tag{5}$$

with $\rho > 0$, is minimized for an increasing sequence of values of ρ, where the penalty part of expression (5) P(T), is defined as

$$P(T) = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \sum_{k=2}^{m} \sum_{l=1}^{k-1} (u_{ijk\ell} - v_{ijk\ell})^2 \tag{6}$$

where:

$$u_{ijk\ell} = \max (t_{i\ell}, t_{ik}, t_{jk}, t_{j\ell}), \tag{7}$$

and

$$v_{ijk\ell} = \begin{cases} \max (t_{i\ell}, t_{jk}, t_{j\ell}) & \text{if } u_{ijk\ell} = t_{ik} \\ \max (t_{ik}, t_{jk}, t_{j\ell}) & \text{if } u_{ijk\ell} = t_{i\ell} \\ \max (t_{i\ell}, t_{ik}, t_{j\ell}) & \text{if } u_{ijk\ell} = t_{jk} \\ \max (t_{i\ell}, t_{ik}, t_{jk}) & \text{if } u_{ijk\ell} = t_{j\ell} \end{cases} \tag{8}$$

The specific steps of the penalty function algorithm are given in De Soete, DeSarbo, Furnas and Carroll (Note 4).

b) Construct a square $(n+m)$ by $(n+m)$ matrix D which satisfies the ordinary one-class ultrametric inequality. $D = ((d_{ab}))$, for $a,b = 1,\ldots,n+m$ is symmetric ($d_{ab} = d_{ba}$) and is defined for $a \neq b$. Because of symmetry, we need only define d_{ab} for $a > b$. D can be thought of as the matrix having T as the $n \times m$ submatrix consisting of the last n rows and the first m columns. The problem is to fill in the (lower half of the symmetric) $m \times m$ and $n \times n$

submatrices comprising the first m rows and columns and the last n rows and columns respectively. This is accomplished by use of the following equations (Furnas, Note 3):

$$d_{ab} = \begin{cases} t_{(a-m)b} & \text{if } m+1 \leqslant a \leqslant m+n \\ & \text{and } 1 \leqslant b \leqslant m \\ \min\limits_{i=1\ldots n}[\max\,(t_{ia},t_{ib})] & \text{if } 1 \leqslant a \leqslant m \\ & \text{and } 1 \leqslant b \leqslant m \\ \min\limits_{k=1\ldots m}[\max\,(t_{(a-m)k},t_{(b-m)k})] & \text{if } m+1 \leqslant a \leqslant m+n \\ & \text{and } m+1 \leqslant b \leqslant m+n. \end{cases} \tag{9}$$

If necessary, a positive constant is added to the d_{ab} so that they satisfy the triangle inequality.

c) Using standard hierarchical clustering methods (see Johnson, (1967)), the ultrametric tree representation of both row and column elements is obtained from D.

2. For Estimating a Path Length or Additive Tree

As discussed in Carroll (1976), Carroll and Pruzansky (1980), Carroll, Clark and DeSarbo (Note 1), based on the work of Farris (1972) and Hartigan (1975), given an ultrametric tree, it is possible to convert it into an additive tree by adding a trivial "star" or "bush" tree (i.e., an additive tree having only one interior node) to it. The algorithm here is thus based on the fact that any set of path length tree distances can be decomposed into a set of ultrametric distances plus a set of additive constants for each of the row and column elements. The numerical problem can be stated as:

$$\text{Min } \{L(T,r_i,c_k) = \sum_{i=1}^{n} \sum_{k=1}^{m} (\delta_{ik}-t_{ik}-r_i-c_k)^2\}, \tag{10}$$

subject to the condition that T satisfies the two-class ultrametric inequality. Once T is estimated via the algorithm described earlier, the additive constants are estimated in closed form via:

$$\begin{aligned} \hat{r}_i &= \frac{\sum\limits_{k=1}^{m} (\delta_{ik}-t_{ik})}{m} - \frac{1}{2}\,\frac{\sum\limits_{j=1}^{n}\sum\limits_{k=1}^{m} (\delta_{jk}-t_{jk})}{nm} \\ \hat{c}_k &= \frac{\sum\limits_{i=1}^{n} (\delta_{ik}-t_{ik})}{m} - \frac{1}{2}\,\frac{\sum\limits_{i=1}^{n}\sum\limits_{\ell=1}^{m} (\delta_{i\ell}-t_{i\ell})}{nm}, \end{aligned} \tag{11}$$

as generalized from Carroll and Pruzansky (1980). Note that this computation of the $\hat{r}_i$ and $\hat{c}_k$ gives them both zero means. This is the exact convention used for fixing the indeterminate constant in the relative lengths of the terminal branches of the row and column objects. Once the $\hat{r}_i$ and $\hat{c}_k$ constants are estimated, the algorithm cycles back to the ultrametric tree estimation phase in estimating T given $\hat{r}_i$ and $\hat{c}_k$. This alternating least-squares procedure (Wold, (1966)) continues cycling back and forth over these two major phases until convergence in the loss function and/or t_{ik} values is reached. Once the final estimates of T and the r_i and c_k are obtained, D is reconstructed from T using the method in expression (9), and the appropriate additive constants are added. The additive tree is recovered from this tree (see Dobson (1974)). This can be done by simply converting the ultrametric tree into an additive tree, by defining the length of every branch to be the difference in height values of the two nodes connected by that branch (thus defining the heights of terminal nodes to be zero), and then adding the constants $\hat{r}_i$ and $\hat{c}_k$ to the lengths of the "leaves" of the tree (the branches connecting the terminal nodes to the first nonterminal or internal node). Because of the indeterminacy mentioned earlier, however, an additive constant can be added to the leaves corresponding to the elements of one class and that same constant subtracted from those of the other set. Finally, if desired, the root node of the ultrametric tree can be removed since that node is redundant for a path length or additive tree, where the two branches issuing from that root node (assuming the tree starts at that node), being replaced by a single branch whose length is the sum of the lengths of the two. Recall that a path length or additive tree is, in a fundamental sense, unrooted; also that the ultrametric tree corresponding to it is not unique - so that the root is highly arbitrary, and could in fact be placed between any two nodes or even at any node of the additive tree. However, for most purposes, it is convenient to represent the path length tree as a hierarchical tree, implying a root. While the root is not unique, the present fitting procedure would tend to place it at the root of the "dominant" or best fitting ultrametric tree, which seems intuitively a reasonable choice. Therefore, in many cases, it may be desirable to retain the root node and the lengths of its two branches, despite the fact that it is redundant.

4. APPLICATION

Bass, Pessemier, and Lehmann (1972) conducted an experiment with 280 students and secretaries. These subjects were required to select a 12-ounce can of soft drink four days a week for three weeks from among the following brands: Coke, 7-Up, Tab, Like, Pepsi, Sprite, Diet Pepsi, and Fresca. Table 1 presents the asymmetric brand switching matrix for the two periods measured.

		Period							
		$t + 1$							
		Coke	**7-Up**	**Tab**	**Like**	**Pepsi**	**Sprite**	**Diet Pepsi**	**Fresca**
	Coke	.612	.107	.010	.033	.134	.055	.013	.036
	7-Up	.186	.448	.005	.064	.140	.099	.012	.046
Period	**Tab**	.080	.120	.160	.360	.080	.040	.080	.080
	Like	.087	.152	.087	.152	.239	.043	.131	.109
	Pepsi	.177	.132	.008	.030	.515	.075	.026	.037
	Sprite	.114	.185	.029	.071	.157	.329	.029	.086
	Diet Pepsi	.093	.047	.186	.093	.116	.093	.256	.116
	Fresca	.226	.093	.053	.107	.147	.107	.067	.200

Table 1

Brand switching matrix for eight brands of soft drinks.

DeSarbo (1982) applied his GENNCLUS procedure to a normalized version of this data set and found three sets of overlapping clusters for the row and column soft drinks: a cola cluster, a diet drink cluster, and a lemon-lime cluster.

We applied our methodology to fit an ultrametric tree to the data in Table 1 and the tree is shown in Figure 3. The tree accounts for 87.6% of the variance of the data. Note, in Figure 3, the underlined soft drinks represent the row objects and the non-underlined soft drinks represent the column objects. One immediately notes the differences in the heights at which different row brands of soft drinks are joined with their same respective column counterparts. For example, Coke_t is joined with Coke_{t+1} at a low height in the tree indicating a propensity of people in the study to repeat trial of Coke - an indicator of strong brand loyalty. Pepsi and 7-Up also display strong brand loyalty in a similar manner. However, some soft drinks (e.g. Tab and Like) do not have such strong brand loyalty, as

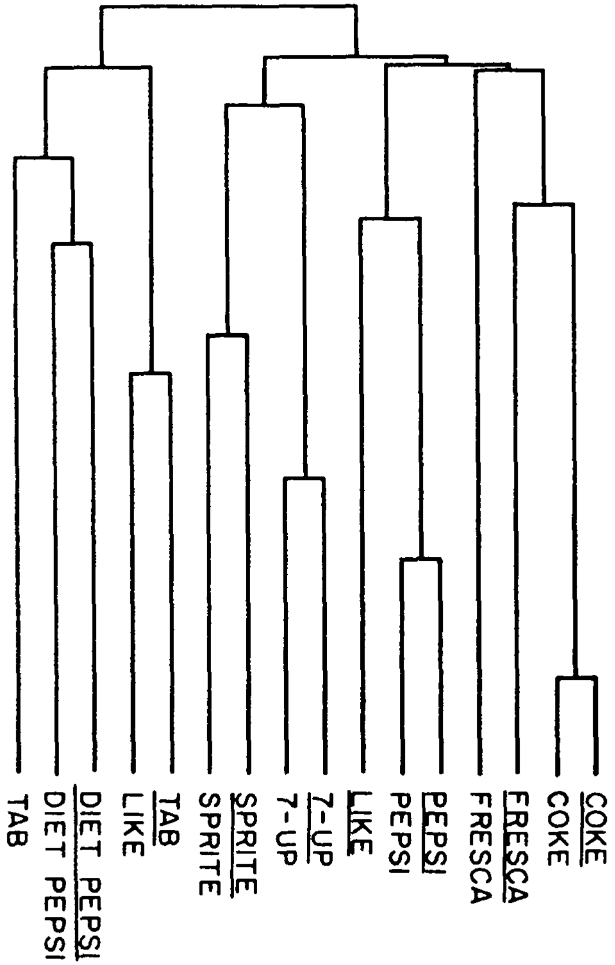

TAB
DIET PEPSI
DIET PEPSI
LIKE
TAB
SPRITE
SPRITE
7-UP
7-UP
LIKE
PEPSI
PEPSI
FRESCA
FRESCA
COKE
COKE

Figure 3. Ultrametric Tree Representation of the Soft Drink Brand Switching Data.

demonstrated by the larger height at which their row and column objects join in Figure 3, as well as their relatively small main diagonal elements in Table 1. Thus, the height at which the *same* row and column object join is *inversely* related to the brand loyalty or main diagonal elements of Table 1.

In one cluster of the tree, we find Tab$_{t+1}$, Diet Pepsi$_{t}$, Diet Pepsi$_{t+1}$and close to this, Like$_{t+1}$ and Tab$_t$. These comprise a diet soft drink cluster. In another cluster, Sprite$_{t+1}$, Sprite$_t$, and close to this, 7-Up$_{t+1}$ and 7-Up$_t$, comprise a lemon-lime cluster of soft drinks.

Finally, one finds Like$_t$, Pepsi$_t$, Pepsi$_{t+1}$ in one group and Fresca$_t$, Fresca$_{t+1}$, Coke$_t$ and Coke$_{t+1}$ in a group next to it. It is a mixed cluster with non-diet colas and Fresca$_t$, Fresca$_{t+1}$ and Like$_t$. Fresca is placed here because of its substantial switching to Coke$_{t+1}$, Pepsi$_{t+1}$, and itself. Like$_t$ is placed here because of the high switching probability to Pepsi$_{t+1}$. Thus, the tree graphically represents the brand switching phenomena presented in Table 1. Here, too, switching can be *roughly* categorized according to non-diet colas, lemon lime drinks, and diet drinks.

5. DISCUSSION

We have presented a methodological description of the model and algorithm which provides an ultrametric or additive tree representation for row and column objects in a two-way, two mode rectangular proximity matrix. The example presented in the previous section illustrates how the procedure can render insight into the structure of a rectangular proximity matrix.

There are a number of possible extensions for this procedure. One obvious extension is to redesign the algorithm(s) in order to estimate multiple tree structures as done in Carroll and Pruzansky (1980) and Carroll, Clark, and DeSarbo (Note 1). This proposed model can be formally expressed in the ultrametric case as:

$$\Delta \cong T_1 + T_2 + \ldots + T_R, \tag{12}$$

where T_1 through T_R arise from R different ultrametric tree structures. In the case of additive trees, expression (12) becomes:

$$\Delta \cong T_1 + T_2 + \ldots + T_R + A, \tag{13}$$

where $A = ((a_{ij}))$ and $a_{ij} = r_i + c_j$ from expression (11)).

Similarly, a procedure can be generalized to accomodate "hybrid models" where both discrete tree structures and continuous spatial dimensions are estimated. Here, for the ultrametric case, one can express this as:

$$\Delta \cong T_1 + T_2 + \ldots + T_R + H, \tag{14}$$

where H is a proximity matrix generated from, say, an S-dimensional non-symmetric multidimensional scaling procedure (e.g., Harshman, Note 5). One merely has to add A from expression (13) to expression (14) to express the hybrid model for the additive tree case.

Another logical extension, and one that is currently being pursued by the authors, is to generalize this procedure to handle three-way rectangular proximities arrays where, as in INDTREES (Carroll, Clark, and DeSarbo (Note 1)), a common tree topology is estimated across slices of the array, allowing for differential branch lengths or node heights.

REFERENCE NOTES

[1] Carroll, J.D., Clark, L.A., and DeSarbo, W.S., The representation of three-way proximities data by single and multiple tree structure models, Unpublished Paper, Bell Laboratories, Murray Hill, N.J. (1983).

[2] De Soete, G., A least-squares algorithm for fitting additive trees to proximity data, Unpublished Paper, University of Ghent, Belgium (1982) in press, Psychometrika (1983).

[3] Furnas, G.W., Objects and their features: The metric representation of two class data, Unpublished Doctoral Dissertation, Stanford University (1980).

[4] De Soete, G., DeSarbo, W.S., Furnas, G.W., and Carroll, J.D., The estimation of ultrametric and path length trees from rectangular proximity data, Unpublished Paper, Bell Laboratories, Murray Hill, N.J. (1983).

[5] Harshman, R, Models for analysis of asymmetrical relationships among *N* objects or stimuli, Unpublished Paper, University of Western Ontario, Canada (1978).

[6] Sarle, W.S., personal communication (1982).

REFERENCES

[1] Bass, F.M., Pessimier, E.A., and Lehmann, D.R., An experimental study of relationships between attitudes, brand preference, and choice, Behavioral Science, 17 (1972), 532-541.

[2] Carroll, J.D., Spatial, non-spatial and hybrid models for scaling, Psychometrika, 41 (1976), 439-463.

[3] Carroll, J.D., and Chang, J.J., A method for fitting a class of hierarchical tree structure models to dissimilarities data, and its application to some body parts data of Miller's. In Proceedings of the 81st

Annual Convention of the Am. Psych. Ass., Vol. 8 (1973), 1097-1098.

[4] Carroll, J.D. and Pruzansky, S., Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares. Paper presented at the U.S.-Japan Seminar on Multidimensional scaling, University of California at San Diego, La Jolla, California, August 20-24 (1975).

[5] Carroll, J.D. and Pruzansky, S., Discrete and hybrid scaling models, in E.D. Lantermann and H. Feger (eds.), Similarity and choice, Bern, Hans Huber (1980).

[6] Coombs, C.H., A Theory of Data, New York, Wiley (1964).

[7] Cunningham, J.P., Free trees and bidirectional trees as a representation of psychological distance, Journal of Mathematical Psychology, 17 (1978), 165-188.

[8] DeSarbo, W.S., GENNCLUS: New models for general nonhierarchical clustering analysis, Psychometrika, 47 (1982), 449-475.

[9] Dobson, A.G., Unrooted trees for numerical taxonomy, Journal of Applied Probability, 11 (1974), 32-42.

[10] Farris, J.S., Estimating phylogenetic trees from distance matrices, American Naturalist, 106 (1972), 645-668.

[11] Hartigan, J.A., Clustering Algorithms, New York, Wiley (1975).

[12] Hartigan, J.A., Modal blocks in definition of west coast mammals, Systematic Zoology, 25 (1976), 149-160.

[13] Hartigan, J.A. and Engelman, L., Block clustering, in the BMDP Statistical Software Manual, revised printing, Univ. of Cal. Press, Berkeley CA (1983).

[14] Hartigan, J.A., Representation of similarity matrices by trees, Journal of the American Statistical Association, 62 (1967), 1140-1158.

[15] Johnson, S.C., Hierarchical clustering schemes, Psychometrika, 32 (1967), 241-254.

[16] McCormick, W.T., Schweizer, P.J., and White, T.W., Problem decomposition and data reorganization by a clustering technique, Operations Research, 20 (1972), 993-1009.

[17] Miller, G.A., and Nicely, P.E., An analysis of perceptual confusions among some English consonants, Journal of the Acoustical Society of America, 27 (1955), 338-352.

[18] Pruzansky, S., Tversky, A., and Carroll, J.D., Spatial versus tree representations of proximity data, Psychometrika, 47 (1982), 3-24.

[19] Rao, S.S., Optimization: Theory and applications, New York, Wiley (1979).

[20] Sattath, S., and Tversky, A., Additive similarity trees, Psychometrika, 42 (1977), 319-345.

[21] Shepard, R.N., and Arabie, P., Additive clustering: Representation of similarities as combination of discrete overlapping properties, Psychological Review, 86 (1979), 87-123.

[22] Sonquist, J.A., Multivariate model building: the validation of a search strategy, Institute for social research, University of Michigan, Ann Arbor, Michigan (1971).

[23] Tryon, R.C., and Bailey, D.E., Cluster analysis, New York, Mc Graw-Hill (1970).

[24] Wold, H., Estimation of principal components and related models by iterative least squares, In P.R. Krishnaiah (ed.), Multivariate Analysis, New York, Academic Press (1966).

Address correspondence to: Geert De Soete
Department of Psychology
University of Ghent
Henri Dunantlaan 2
B-9000 Ghent
Belgium

G. De Soete is "Aspirant" of the Belgian "Nationaal Fonds voor Wetenschappelijk Onderzoek". W.S. DeSarbo, G.W. Furnas, and J.D. Carroll are all Members of Technical Staff at the Bell Laboratories in Murray Hill, New Jersey.

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

WEAK AND STRONG MODELS IN ORDER TO DETECT AND MEASURE POVERTY

Paul Dickes

University of Nancy

The paper presents the application of cluster-analytic methods (weak model) and the logistic model of Rasch (strong model) on a list of dichotomized items covering the living conditions of families with children in five different countries of the European Community. Cluster-analytic methods reveal the presence of three typological groups: one of them has practically no disadvantages. The other two are marked by many disadvantages and contain, on the one hand, single-parent families, on the other hand, families of foreign origin. The Rasch model shows that a subset of items, belonging to diverse living conditions, are model-conform. Thus the measurement of poverty through the accumulation of disadvantages has a sense. The typological groups characterized by many disadvantages are to be considered as families at risk of poverty.

1. INTRODUCTION

The distinction between weak and strong models in the methodology of social sciences became more obvious as far as latent trait models (Lord and Novick (1968)) and methods of scaling (Van der Ven (1980)) were developed.

Strong models consist of propositions, derived from postulates or axioms, and leading to empirical operations through which assumptions can be falsified. Weak models are those for which falsificability is reduced. The classical test theory is considered as weak, whereas the modern test theory or the theories of latent traits are generally recognized as belonging to strong models (Fischer (1974); Kristof (1983)).

A second distinction, partially linked with the former one, is attached to the difference between confirmatory and exploratory level of analysis. The confirmatory approach, best exemplified by Joreskog (1974), allows the analyst to utilize theoretical assumptions about data for the selection of parameters, and to test the fit between theoretical and empirical results. The exploratory approach, largely used in the multivariate exploitation of data (such as the factor analysis, the cluster analysis, the multiple regression, etc.), makes no direct assumptions about the structure of these data. The analyst tries to reduce the complexity and multiplicity of data to more common aspects. The exploratory analysis is mainly used as a data-reduction technique.

Strong models and confirmatory analysis satisfy more to epistemological considerations than the application of weak methods. They belong to the hypothetico-deductive process of scientific progress and are singled out by the possibilities of falsification they offer , which are considered crucial in the formulation and demonstration of scientific laws (Popper (1968)). With them an attempt is made to reach the standards of formalized sciences as are the natural sciences.

The interest of exploratory methods lies in the development of hypothetico-inductive processes of scientific discovery. Confronted with a large amount of data with unknown structure, the researcher has the possibility, with these methods, to detect some regularities leading to hypotheses which then are to be confirmed.

The objective of the present paper is to demonstrate the complementarity of the two kinds of approaches in a specific field of social research concerned with the study of poverty. Cluster analytic-methods and latent trait models were applied to a same data matrix, described further on. The results of these two, rather antagonistic approaches, lead to a better understanding of the phenomena, and this would not have been reached without the use of the two methodological instruments.

2. CONDITIONS OF THE RESEARCH

The following research project belonged to the European Program to Combat Poverty which took place in the late seventies (Schaber et al. (1980)). The objectives of the research were to detect and study persistent poverty in seven regions of five Member States (France, Belgium, Germany, Netherlands and Luxemburg).

The investigated population consisted of all families who had a child born in 1970. This rather homogeneous group was chosen in order to study persistency of poverty from one generation to the other.

In each region 200 families were randomly selected from exhaustive lists of population. The total sample consisted of 1400 families. Two stages were foreseen for the research. The first one stretched over the first months of 1979: the 1400 families were interviewed with a 20-30 minutes' questionnaire in order to gain a general knowledge of their living conditions. The purpose of this rather instrumental stage was to discover families at risk of poverty and to develop instruments to measure it. It was a preparatory stage to the second one, which started one year later. There, intensive interviews were performed, the original sample being oversampled with families at risk of poverty. The analysis and the results we present here are based solely on the first-stage data.

3. RATIONALE AND DATA MATRIX

Income is used as the privileged criteria in most researches, mainly anglo-saxon, to measure poverty. Although one can deduce out of the income, especially in western industrialized countries, many aspects of living conditions, it appears as particularly unsuitable in comparative research. Comparisons between countries with diverse economical and political conditions are quite difficult and often inappropriate. Even within the same region or country, if income can be chosen as main poverty criteria, it still remains difficult to study the actual effect of income on poverty.

The start position of this research is at a first open and not constraining. Poverty concerns different domains of living conditions. Thus we selected 86 variables out of the scientific literature and/or practical experiences with poor people, that were considered as characteristics or correlates of poverty. These variables cover diverse domains of the living conditions of populations as for example: family structure, nationality, ethnic origin, job situation, educational background, economic situation, housing commodities, health conditions, cultural and relational practices. Most variables were situational but some were subjective as for example feelings of well-being, feelings of deprivation, etc.

The answers to the questions were dichotomized. In this way a list of 86 binary items actualized the disadvantages of each family. The final data matrix consisted of a vector of 86 items and a vector of 1400 families (persons). The coded answers to each item were either:

1: if the family is marked by a given disadvantage, or
0: if the family is not marked by the disadvantage.

The hybrid nature of measurement levels directed our choice of binary response structure. Some original variables are nominal, others ordinal, of equal interval or even rational. The transformation of higher level scales to a lower level one has no direct mathematical limitation, but the reverse is not necessarily true.

4. TYPOLOGICAL IDENTIFICATION OF GROUPS AT RISK OF POVERTY

The utilization of cluster-analytic procedures doesn't require any theoretical assumption about data. All the subjects in the sample are situated in a space which has as many dimensions as there are items.

Out of the relations between the persons it is possible that clusters emerge, characterized by the fact that the similarities within the groups are greater than the similarities between the groups.

The cluster-analytic techniques are based on weak models in the sense that they have no algorithm usable for discarding items or persons. The obtained results are exclusively dependent on items and subjects submitted to analysis. The approach is an exploratory one, in the sense that unknown structure or order have to be discovered and interpreted.

We consider that the cluster-analysis reaches the expectations if:

a) the typological attribution is stable (or reliable), and
b) if at least one of the identified groups can be recognized as marked by more disadvantages than the other(s).

The purpose of the analysis is to identify group(s) at risk of poverty, without presupposing an explicit definition of what poverty or poverties really is or are. We would succeed if the knowledge we gained out of the sample could be applied to the population. It would then be possible to detect families at risk of poverty and to develop oversampling schemes.

The reliability of the classification was estimated through two types of procedures. First, the original sample was partitioned into odd and even subsamples in order to develop crossvalidation calculations. Then, three different cluster-analytic methods were applied to each subsample. Thus invariance of results was expected over samples and methods.

There exists a large range of cluster-analytic methods, among which a researcher has to choose appropriate ones to his specific purposes. For this research only hierarchical clustering techniques were used. They are best suited for large data matrix and don't preclude any number of clusters, as other techniques do.

Three clustering-techniques were used here: the complete linkage method, the Wards method of cluster-analysis, and an associative method.

The first two methods are agglomerative and the last one is divisive. The euclidean distance was employed as dissimilarity coefficient in the agglomerative techniques. The Chi-square was the dissimilarity coefficient of the divisive technique.

The analysis was performed until 8 clusters were obtained. To each of the eight-clusters solutions, reallocation techniques, based on euclidean distance, were applied. The optimal number of clusters was reached when convergence was obtained between the reallocated clusters solutions. This occured at the level of three-clusters solutions for odd and even subsamples of subjects.

Crossvalidation was performed in the following ways:

- for each subsample, the two discriminant functions were calculated for the three clusters.
- these equations were applied across the samples. This means that the discriminant functions of the odd-subsample were applied to the even-subsample as well as the discriminant functions of the even-subsample to the odd-subsample.
- the obtained functions, produced very significant canonical correlations: odd-subsample Rcan = 0.90 and 0.86 and even-subsample Rcan = 0.92 and 0.86. Practically 94% of correct classifications were obtained. 84% and 87% of the groups 2 and 3 were correctly classified. Thus, the groups appeared as largely comparable over the samples.

The three groups can be described as follows:

- the first group contains 73% of the total sample and is characterized by the absence of disadvantages.
- the second group (19% of the total sample) shows an accumulation of disadvantages in many domains as for example: housing, health, income.

There the families are complete (father and mother present and married) and mainly of foreign origin.

- the third group (8% of the total sample) presents the same accumulation of disadvantages as the second one. However, in this group the father is seldom head of the household, the main income comes from social transfer or social assistance; the families are native of the regions where they live and are mainly broken or incomplete homes.

Between these three stable groups, two of them can be recognized as being at risk of poverty. If the following statement is true: poverty appears when families present disadvantages in multiple domains, then the two groups "at risk" must include the "poorest". But they may also include "non-poor", which means that these families are at risk of poverty, but that they don't manifest poverty (not yet or not longer).

At this stage of the research, we didn't know if the accumulation of disadvantages presented by the groups "at risk", had a single and sole signification. At the experimental level, the only possible indication we had, was the following: the two groups "at risk" had to be more heterogeneous than the non-disadvantaged group, and had to present more disadvantages. The typological analysis shows that the life situations of being "at risk of poverty" vary in a qualitative way. Even if we haven't yet worked out a clear and precise definition of poverty, our pre-scientific notions have allowed us to describe typical precarious family situations.

5. OBJECTIVE MEASUREMENT OF POVERTY

In the exploratory stage of data analysis, the accumulation of disadvantages has helped to differentiate groups at "risk" and at "non-risk" of poverty.

A more elaborate hypothesis about data can now be reasonably formalized: poverty is the accumulation of disadvantages in different domains of the living conditions. This hypothesis is in agreement with the traditional research, where poverty is solely defined in income terms. Poverty is considered as a one and a lone continuum. The hypothesis about unidimensionality needs a strong model in order to be tested.

The classical test theory (Gulliksen, (1950)) assumes unidimensionality of the constructs, but is particularly weak in model-testing (Novick (1966); Fischer (1974), Kristof (1983)).

Latent trait models appear much more satisfactory. They postulate the unidimensionality of the construct (latent trait), have explicit parameters about items and persons and lead to propositions from which assumptions could be tested (Fischer (1974)).

Between different latent trait models (Lawley (1943); Lord (1952), (1953): Lord and Novick (1968): Fischer (1974)) the Rasch model (Rasch (1960), (1968)) appears as particularly appropriate to the pursued purpose (Gustaffson (1977)). He is objective, parsimonious (one-parameter model), easily interpretable, has satisfactory algorithms in order to estimate the parameters and is easily testable.

The logic of the model and the procedures to resolve the parameters need not to be presented here. Sufficient literature is available (Rasch (1966); Fischer (1974)).

The data matrix of 86 items and 1400 families constituted the starting point of the analysis. We used a computer program implemented by Gustaffson (1977). The estimation of the parameters was obtained through conditional likelihood algorithms.

In a first step, very strong criteria were applied in order to test the fit between model and data.

Five examples were constructred:

- Sample number 1: total sample
- Sample number 2: typological group 1
- Sample number 3: typological group 2
- Sample number 4: typological group 3
- Sample number 5: typological groups 2 and 3.

From the 86 original items, 11 were recognized as model-conform. The estimation of item-parameters remained practically invariant over the five subsamples. This item-set was considered as "anchor-set". At first glance the retained items covered different domains: health, income, education, etc.

In the second step, the previous selection criteria was lightened. It was expected to find an invariance of items' estimations over the samples number 1, 2 and 5. From the original item list, 31 remained conform to the model. This new set contained the "anchor" items, already tested, but also other items, which had been found as model-conform. The remained items largely covered the different domains of living conditions of the families.

Because the hypothesis of unidimensionality of the 31 items could not be rejected, we concluded that the items of this set measured a sole and same thing. Out of the obvious heterogeneity of the domains to which belonged the items, we concluded that only a reference to poverty could best interpret the homogeneity of the latent trait. This was confirmed by the fact that the "hard core variables", traditionally associated with poverty, as for example income, social transfer, deprivations-, belonged to the retained list.

Thus, our first hypothesis seemed confirmed and poverty appeared as an accumulation of different disadvantages, in different domains of the living conditions. This hypothesis would have been rejected if the retained items belonged to one specific domain (like housing or health). In this case the latent continuum would have only revealed one specific accumulation of disadvantages. If we had been able to find two or more continua of this kind, we could have talked about poverties but not about poverty. In our case however, we measured a latent continuum, interpreted as describing poverty.

Each item has a specific position on this trait. The higher the position (estimation of item-parameter), the harder is the poverty it reveals. Items with low values point out to minor disadvantages (as for example overcrowding, having no books in the house, etc.). Items with high values reveal serious poverty situations (as for example the dependance on social aid, being sub-tenant, etc.).

Each person has also a specific position on this same continuum. Rasch-model has proprieties that allow to make the summation of all the disadvantages. In this way to each person corresponds a "value", which is a sufficient statistic to make person's estimate on the scale. Thus the more disadvantages a family has, poorer it is and the higher the value it reaches.

The measurement is objective in the sense that 1) the person's parameter is independent of the item-set and 2) the item's parameter is independent of the samples of persons. Thus, the measurement of the person's parameter can be realized with any sub-set of model-conform items and cound be applied to any subsamples of persons.

These properties of the Rasch model will allow several operations, as for example:

- to compare the results of different investigations with the same scale if at least some model-conform items are common;
- to create a bank of items where poverty items would be characterized by their parameter;
- to constitute user-suited scales for specific research purposes;
- to cumulate knowledge from one study to the other, in order to purify or to enlarge the range of the original item-pool;
- to study change.

6. RISK AND MEASUREMENT OF POVERTY

The exploratory analysis of the data matrix in the hypothetico-inductive stage of the research reveals three typological groups. Two of those were analysed as being at risk of poverty. In our particular sample, in the chosen European countries, it appears that disadvantages are more frequent in one-parent families and in families belonging to ethnic groups different from the one of the country they live in. The obtained results are dependent on the nature of the sample: if other populations had been studied, other groups at risk would probably have emerged. These results are in themselves interesting, and allow practical applications to be made. One example could be to detect groups at risk of poverty, in order to oversample them in a next research.

This first analysis doesn't however, say anything about the measurement of poverty. Is one person poorer than the other? Do the disadvantages of one group correspond to the disadvantages of the other group? Can the accumulation of disadvantages, as we observed it in the high-risk groups, be linked to a unique concept? All these questions needed to be answered, so a precise measurement instrument can be made available.

In the confirmatory stage an explicit hypothesis about the data matrix was tested: poverty has a unique dimension and is characterized by the accumulation of disadvantages in different domains. The utilization of a strong model allowed us to confirm the initial hypothesis, and to construct a measuring instrument which had the characteristics of objectivity.

We expect that in other samples or other populations the hypothesis will likewise be confirmed. In this way, the hypothetical statement could be tested again and the measuring instrument improved.

Figure 1 shows the distribution of the typological groups in relation with the number of disadvantages they accumulate.

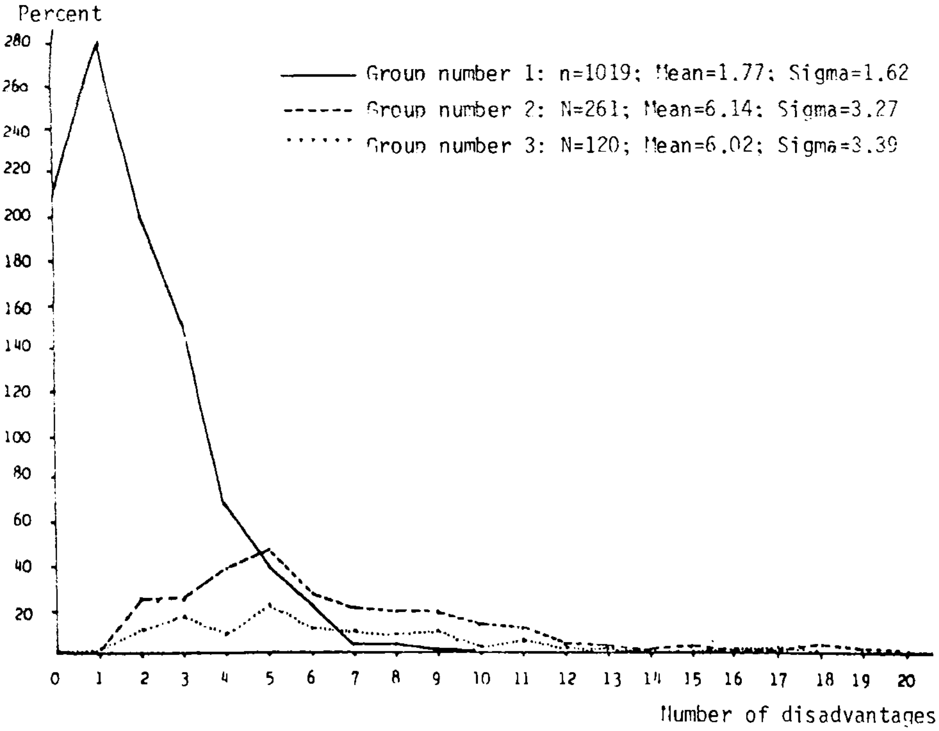


Figure 1: Number of disadvantages and typological groups.

The two groups at risk present more disadvantages and are more heterogeneous than the non-disadvantaged group.

Thus the confirmatory analysis allows an interpretation of the descriptive statements of cluster-analytic techniques: the groups at risk of poverty present more disadvantages and the accumulation of disadvantages is poverty.

REFERENCES

[1] Fischer, G.H., Einführung in die Theorie psychologischer Tests, Bern, Huber (1974).

[2] Gulliksen, H., Theory of mental tests, New York, Wiley (1950).

[3] Gustafsson, J.E., The Rasch model for dichotomous items: theory, application and a computer program, Goteborg, Institute of Education, (1977)

[4] Joreskog, K.G., Analyzing psychological data by structural analysis of covariance matrices, in Krantz, D.H. et al. (Eds.), Contemporary developments in mathematical psychology, Vol. II, San Francisco, Freeman (1974), 1-56.

[5] Kristof, W., Klassische Testtheorie und Testkonstruktion, in Feger, H. & Bredenkamp, J. (Eds.), Messen und Testen, Enzyklopädie der Psychologie, Göttingen, Hogrefe, Bd. 3 (1983), 544-603.

[6] Lawley, D.N., On problems connected with item selection and test construction, Proceedings of the Royal Society of Edinburgh, 61-A (1943), 273-287.

[7] Lord, F.M., A theory of test scores, Psychometric Monographs, 7 (1952).

[8] Lord, F.M., The relation of test scores to the trait underlying the test, Educational and Psychological Measurement, 13 (1953), 517-548.

[9] Lord, F.M., and Novick, M.R., Statistical theories of mental test scores, Reading, Mass., Addison-Wesley (1968).

[10] Novick, M.R., The axioms and principal results of classical test theory, Journal of Mathematical Psychology, 3 (1966), 1-18.

[11] Popper, K.R., The logic of scientific discovery, London, Hutchinson (1968).

[12] Rasch, G., Probabilistic models for some intelligence and attainment tests, Copenhagen, Nielsen and Lydiche (1960).

[13] Rasch, G., A mathematical theory of objectivity and its consequences for model construction, Paper delivered at the European Meeting on Statistics, Econometrics and Management (1968).

[14] Schaber, G et al., Pauvreté persistente- grande région, Projet 28, Commission des Communautés Européennes, Groupe d'étude pour les problèmes de la pauvreté (1980).

[15] Van der Ven, A., Einführung in die Skalierung, Bern, Huber (1979).

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

APPLICATIONS OF A BAYESIAN POISSON MODEL FOR MISREADINGS

Margo G.H. Jansen

University of Groningen

Apart from the widely known model for binary scored items, Rasch, has developed several other models for the analysis of achievement test data. The model we consider here is the so-called multiplicative Poisson model for misreadings. This model assumes that the total number of misreadings on a text for a certain individual is approximately Poisson distributed with an intensity parameter μ which depends on the ratio of two other parameters, one pertaining to the ability of the individual and one to the difficulty of the text. A Bayesian version of this model was developed by Owen (1969). In this paper we adopt a slightly different formulation in the tradition of the general approach by Lindley (Lindley and Smith (1972)). The method is illustrated by an empirical example.

1. INTRODUCTION

Apart from the so-called one-parameter logistic model for binary scored items, Rasch has developed several other latent trait models for the analysis of achievement test data. Among these are the models for oral reading speed and for misreadings in a text. These models are described in the monograph "Probabilistic models for some intelligence and attainment tests", Rasch (1960), and summarized by Lord and Novick (1968).

Compared to the logistic model these models have attracted much less attention. As a result these models are much less developed and practical applications apart from those of Rasch himself are rare. This seems unfortunate because tests that would lend themselves excellently to analyses based on Poisson process models are widely used in educational practice.

Bayesian versions of the Poisson model for misreadings were already developed in 1969 by Owen. Related models were developed by Van der Ven (1969).

2. THE POISSON MODEL FOR MISREADINGS

In contrast to the one-parameter logistic model, the Poisson model for misreadings is a latent trait model for tests rather than items. Rasch however, begins his development of the latter with the assumption that a text to be read can be considered as a succession of mutually independent Bernouilli trials, corresponding to the words in the text, where the chances of an error are relatively small. A second, but extremely important assumption is that the probability of a misreading is a ratio of two factors, one pertaining to the ability of the pupil and one pertaining to the difficulty of the words.

Then, given that the error probabilities are indeed small and the number of words in the text (= number of items) is large, the distribution of the total number of errors in the text (= test score) may be approximated by a Poisson distribution with intensity parameter equal to the sum of the Bernoulli parameters. Rasch assumed that the error probabilities within a given test and person combination were constant but this is not necessary, a point also made by Lord and Novick, ch. 21 (1968). The intensity parameter has the same multiplicative structure as the parameters on item-level.

In fact, we could take the Poisson model for the total scores as our starting point, with the item-level model as one possible underlying structure. Looking at the examples given by Rasch, it is clear it is the model at the test-level for which we have empirical evidence. The test-level model is completed by the assumption that we have two or more tests for which a Poisson model holds, and that the usual conditions of local independence and independence accross pupils hold.

Formally, consider a sample of N pupils taking k tests, where the distribution of the number of errors made by person i on test j is given by

$$P(Y_{ij} = y_{ij}|\mu_{ij}) = \{\exp(-\mu_{ij})\}\, \mu_{ij}^{\,y_{ij}} / y_{ij}! \qquad (1)$$

We suppose that the intensity parameter μ_{ij} is a ratio of two other parameters, the difficulty δ_j of the test and the ability ξ_i of the person.

$$\mu_{ij} = \delta_j \,/\, \xi_i . \tag{2}$$

The expected number of errors will be higher if the test is more difficult and/or the person is less able and lower if the test is less difficult and/or the person more able.

Obviously neither, the δ_j nor the ξ_i are uniquely determined, for the k tests and N persons. This state of affairs can be remedied by imposing an arbitrary restriction on the set of parameters. Rasch proposes taking the most difficult test as a point of reference by choosing a value of one for the corresponding difficulty parameter. Notice that, because all Poisson parameters have to be nonnegative, setting one difficulty parameter to be positive, implies that all difficulty parameters δ_j and all ability parameters ξ_i are also nonnegative.

In the Bayesian framework, identifiability problems are solved differently, by choosing suitable prior distributions which tell us how to identify the parameters, as we will show in the next section.

From a theoretical point of view, one very desirable characteristic of this model, which it shares with the "ordinary" Rasch-model, is the separability of the parameters. Test parameters can be estimated independently of the person parameters and vice versa. For a discussion of what is also known as "specific objectivity", see Rasch (1980) and Wright and Stone (1979).

3. BAYESIAN VERSIONS OF THE MULTIPLICATIVE POISSON MODEL

A full Bayesian model for the multiplicative Poisson model for misreadings has been developed by Owen (1969). He took the formulation in Eqns (1) and (2) as starting point but adopted another definition of the parameters. The person parameter he used is inversely proportional to its counterpart in Rasch's formulation. This change was brought about to provide a complete symmetry between the person and test parameters. Prior information was specified by choosing mutually independent gamma distributions for test and person parameters, and several posterior densities, including the conditional and marginal posterior densities, were derived.

A similar approach was used by Van der Ven (1969), in order to find suitable strong true-score models for the analysis of time-limit tests. Our own formulation is more similar to the approach of Leonard (1971) and that of Lindley (1970). We adopt the following transformation of the original ability and difficulty parameters

$$\theta_i = -\ln\xi_i \quad \text{and} \quad b_j = -\ln\delta_j$$

and

$$\ln(\mu_{ij}) = \theta_i - b_j \tag{3}$$

The likelihood function in terms of θ and b is given by

$$L(\theta,b) = \exp\,\{\Sigma y_{ij}(\theta_i - b_j) - \Sigma \exp\,(\theta_i - b_j)\} \,/\, \Pi(y_{ij})! \tag{4}$$

Treating $\theta_1,\dots,\theta_N$ respectively $b_1,\dots,b_k$ as exchangeable within each set we assume the following mutually independent marginal prior distributions

$$\theta_i \sim N(\mu_\theta,\phi_\theta) \tag{5}$$

and

$$b_j \sim N(0,\phi_b) \tag{6}$$

where μ_θ, ϕ_θ and ϕ_b are supposed to be known constants.

A more convenient formulation is based on the following. Let

$$\theta_i^* = \theta_i - \mu_\theta . \tag{7}$$

Then the prior distributions can be formulated as follows.

$$\theta_i^* \sim N(0,\phi_\theta) \tag{8}$$

and

$$b_j \sim N(0,\phi_b). \tag{9}$$

The prior variances are assumed known, as in the previous formulation. The mean μ_θ can also be assumed known or be retained as an explicit parameter. In that case we assume that μ_θ is distributed uniformly over the real line. All prior distributions are considered mutually independent. The joint posterior distribution is proportional to

$$p(\theta^*,b,\mu_\theta|Y,\phi_\theta,\phi_b) \propto$$

$$L(\theta^*,b,\mu_\theta|Y) \times$$

$$\phi_\theta^{-\frac{1}{2}n} \exp(-\frac{1}{2} \sum_i \frac{\theta_i^{*2}}{\phi_\theta})\, \phi_b^{-\frac{1}{2}k} \exp(-\frac{1}{2} \sum_j \frac{b_j^2}{\phi_b}) . \tag{10}$$

Estimates for the parameters can be obtained by taking the logarithm of the posterior density $p(\theta^*, b, \mu_\theta | Y)$, by differentiating with respect to μ, θ_i^* and b_j, setting their derivatives equal to zero and solving the resulting equations:

$$\sum_{i,j} y_{ij} - \sum_{i,j} \exp(\theta_i^* + \mu - b_j) = 0, \tag{11}$$

$$-\sum_j y_{ij} + \sum_j \exp(\theta_i^* + \mu - b_j) + \theta_i^* / \phi_\theta = 0, \tag{12}$$

and

$$-\sum_i y_{ij} + \sum_i \exp(\theta_i^* + \mu - b_j) - b_j / \phi_b = 0. \tag{13}$$

As is obvious from equations (11) through (13) an iterative procedure is required to solve them. Under appropriate conditions, the result will be the mode of the joint posterior density.

4. EXAMPLES

The first example is based on an artificial data set consisting of the responses of 100 individuals on 3 tests. The abilities were randomly selected from a standard normal distribution. For the tests, we choose difficulties of resp. $b_1 = -0.5$ $b_2 = 0.0$ and $b_3 = +0.5$. This corresponds to an error probability of resp. 3, 2 and 1 percent for a "standard" person with ability $\theta = 0$, given a test length of 50 items. The prior variances were assumed to be equal to a value of 1.

The results can be found in table 1. Both the Bayesian and the ordinary maximum likelihood estimates for the parameters θ and b ($-\ln(\xi)$ and $-\ln(\delta)$) are shown.

The Bayesian estimates show the typical regression effect compared to the maximum likelihood estimates. The regression effect is more marked for the ability than for the difficulty estimates, due to the asymmetry in the sample information (100 persons and 3 tests).

These results are similar to results obtained by using Bayesian methods for estimating the parameters of the better known one-parameter logistic model (Jansen & Lewis (1983)).

Raw Score	Theta Estimate 1)		Test Nr.	Beta Estimate	
	ML	BAY		ML	BAY
0	----	-1.16	1	-0.507	-0.503
1	-1.19	-0.72	2	-0.028	-0.027
2	-0.49	-0.38	3	0.535	0.530
3	-0.09	-0.10			
4	0.20	0.13			
5	0.42	0.33			
6	0.61	0.50			
7	0.76	0.64			
8	0.89	0.77			
9	1.01	0.89			
10	1.12	1.00			
11	1.21	1.09			
12	1.30	1.18			
13	1.38	1.27			
14	1.45	1.34			
15	1.52	1.41			
	etc.				

1) $\theta = \theta^* + \mu_\theta$

Table 1

Maximum Likelihood and Bayesian Estimates for the Person and Test Parameters of the Artificial Data Set (N = 100, k = 3).

Another interesting feature is that, with the Bayesian approach, we also have finite estimates for persons with a perfect score (error score of zero). This is important since, with easy tests, a relatively large proportion of the sample consists of just these persons.

For the second example we used an empirically obtained data set. This data set consisted of the responses of pupils on a spelling test that was administered twice under different format conditions. The parameter estimates for the person and tests can be found in Table 2, on the next page.

The results in Table 2 are comparable to those in the first Table.

Raw Score	Theta Estimate		Test Nr.	Beta Estimate	
	ML	BAY		ML	BAY
0	----	-0.41	1	0.116	0.115
1	-0.70	-0.02	2	-0.116	-0.115
2	-0.01	0.28			
3	0.40	0.53			
4	0.69	0.73			
5	0.91	0.91			
6	1.09	1.07			
7	1.25	1.21			
8	1.38	1.33			
9	1.50	1.44			
10	1.60	1.54			
11	1.70	1.63			
12	1.79	1.72			
13	1.87	1.80			
14	1.94	1.87			
15	2.01	1.94			
	etc.				

Table 2

Maximum Likelihood and Bayesian Estimates for the Person and Test Parameters of the Data Obtained by Giving the Same Test under Two Different Format Conditions (N = 716, K = 2).

5. DISCUSSION

The results show that Bayesian methods are potentially useful for estimating the parameters of the multiplicative Poisson model. More research however, is necessary to substantiate this claim.

REFERENCES

[1] Andersen, E.B., Discrete statistical models with social science applications, North Holland, Amsterdam (1980).

[2] Leonard, T., Bayesian methods for multinomial data, ACT Technical Bulletin 4, The American College Testing Program, Iowa City (1971).

[3] Lindley, D.V., A Bayesian solution for some educational prediction problems, III, Research Bulletin RB-70-33, ETS, Princeton (1970).

[4] Lindley, D.V., & Smith, A.F.M., Bayesian estimates for the linear model, Journal of the Royal Statistical Society (Ser. B), 34 (1972), 1-41.

[5] Lord, F.M., & Novick, M.R., Statistical theories of mental test scores, Addison-Wesley, Reading, Mass. (1968).

[6] Rasch, G., Probabilistic models for some intelligence and attainment tests, Copenhagen (1960), (Reprinted by the University of Chicago Press (1980)).

[7] Owen, R.J., A Bayesian analysis of Rasch's multiplicative Poisson model for misreadings, Res. Bull. E.T.S., Princeton (1969).

[8] Ven, A.H.G.S. van der, The Binomial error model applied to time-limit tests, Unpublished doctoral dissertation, Nijmegen (1969).

[9] Wright, B.D., & Stone, M.H., Best test design, MESA Press, Chicago (1979).

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

THE POLYCHOTOMOUS RASCH MODEL AND DICHOTOMIZATION OF GRADED RESPONSES

Paul G.W. Jansen
Rijks Psychologische Dienst, Den Haag,

and

Edw. E. Roskam
Department of Psychology, University of Nijmegen

This paper discusses the compatibility of the polychotomous Rasch model with dichotomization of the response continuum. It is argued that in the case of graded responses, the response categories presented to the subject are essentially an arbitrary polychotomization of the response continuum, ranging e.g. from total rejection or disagreement to total acceptance or agreement of an item or statement, and should not interfere with subject measurement.

A distinction is made between polychotomization or dichotomization 'before the fact', i.e. in constructing the response format, and polycho- or dichotomization 'after the fact', e.g. in dichotomizing existing polychotomous data.

It is shown that, at least in case of dichotomization 'after the fact' the polychotomous Rasch model is not compatible with such dichotomization. If data satisfy the unidimensional polychotomous Rasch model, they will not do so after dichotomization, unless a rather special condition of the model parameters is met. In so far as dichotomization 'before the fact' is not essentially different from dichotomization 'after the fact', the value of the unidimensional polychotomous Rasch model is questionable.

The impact of our conclusion on related models is briefly discussed.

1. INTRODUCTION

Various models exist for the representation and analysis of data where the response format consists of two or more alternatives, categories or options. These models typically concern attitude questionnaires, but may also apply to aptitude items.

Among the probabilistic models, the multidimensional polychotomous Rasch model (Rasch (1961)) appears to be the only model for items with multiple response options, which is in a sense multidimensional (we will comment later on the multidimensional character of this model).

Unidimensional latent trait models represent the subject by a single scalar parameter. For polychotomous data, there are the Likert((1932), cf. Roskam (1983))model, the unidimensional polychotomous Rasch model, and Samejima's (1969) polychotomous logistic difference model. Related to the unidimensional polychotomous Rasch model are Andrich's (1978) Rating Rasch model and Masters' (1982) Partial Credit model. It should be noted that the Likert model is not a probabilistic response model, but actually a polychotomous Guttman scale model with the added assumption that the subjects' parameters are normally distributed in the population of subjects.

The unidimensional models are not necessarily meant to be also models for graded responses, but nothing appears to prevent one from applying them to graded response data, as usually obtained from attitude questionnaires, where response options are typically of the 'degree of agreement' kind. An example of the application of the unidimensional polychotomous Rasch model to attitude data is given by Andersen (1973).

The mere fact that a model does not appear to prevent its application to graded responses does not mean that it is compatible with psychological assumptions and/or methodological requirements which apply to a graded response format. This paper addresses itself to the compatibility of the polychotomous Rasch model with such assumptions and/or requirements in the case of graded responses. In particular, we address ourselves to their compatibility with dichotomizing the responses, either in constructing the response format, or after having collected multiple choice data.

1.1. GRADED RESPONSES

We assume that we have a priori knowledge concerning the ordering of response categories or options whenever their phrasing contains at least ordinal quantifiers such as 'strongly','moderately','sometimes','not at all', and the like, or when the response categories are marked by symbols having a natural semantic order, like 'a', 'b', 'c',..., or '1', '2', '3',... . Obviously, response options like "don't know" do not naturally fit into an ordering of response categories.

The general principle of unidimensional graded response models for attitude questionnaires can be expressed as follows. It is assumed that subjects can be represented at least ordinally as points in a latent unidimensional continuum. With a given item, the response categories correspond to a partitioning of that continuum, meaning that, in a manner of speaking, the subject locates himself on that continuum by picking a response option according to where he stands. The following is an example:

I am able to jump higher than 1.60 meter:

- certainly not
- probably not
- just about yes or no
- probably yes
- certainly yes

Each item cum response category expresses a degree of positive judgment towards its object. This is most clearly so in the deterministic polychotomous Guttman scale model, as presented e.g. by Coombs (1964). The same concept of categories partitioning a latent space is also central to multidimensional scalogram analysis (Guttman (1967); Lingoes (1968)).

In deterministic models, the correspondence between response categories and segments of the latent continuum is one-to-one. This is obviously not the case in probabilistic models. Without speculating on the nature of the stochastic mechanism underlying a non-deterministic response structure,we can still adopt the concept of response continuum which is partitioned into successive categories, such that the probability of a subject's picking a certain response option is a probabilistic function of his position on the latent continuum. In fact, we assume that the subject's expressed response corresponds to an interval on a response continuum which is itself imperfectly correlated with the latent continuum.

Here, like in the deterministic case, each item cum response category is an index of the degree of the subject's positive attitude towards some object, but in the probabilistic case this index is imperfect. A probabilistic response model states, in an explicit algebraic form, the probability that the subject will pick a particular response option, given his latent trait position, and given certain parameters specific to the item and its response categories.

We consider the response categories as intervals along a response continuum, or, as a polychotomization of that continuum. In its simplest form, we have a dichotomization into, e.g. 'yes, agree', and 'no, disagree'.

1.2. POLYCHO- OR DICHOTOMIZATION BEFORE THE FACT OR AFTER

For the present discussion we must distinguish polychotomization (or its extreme: dichotomization) before the fact, from polycho-dichotomization after the fact. In constructing the response options, the response continuum is partitioned into successive intervals before the fact, i.e. before the responses are actually given and collected. If, on the other hand, we combine different responses into a smaller set of categories at the time of data analysis, we (re-)partition the response continuum after the fact, i.e. after the responses are actually given and collected. The latter is common practice when graded responses are dichotomized after the data have been collected, e.g. in order to simplify their analysis.

Keeping this distinction in mind, the question arises whether or not the polycho- or dichotomization of responses interferes with subject measurement. The problem of non-interference 'before the fact' refers to the effect of the structure of graded response alternatives on subject measurement. It would seem only reasonable that it does not matter whether the subject must choose his response from one set of graded responses or from another set of graded responses (both pertaining to the same response continuum), at least not so that that would interfere with whatever it is that is being measured. In particular, this concerns response formats which differ only with respect to the quantifiers used to express degrees of e.g. acceptance, agreement, and the like. Exchanging one set of quantifiers in the response format for another set of quantifiers should be immaterial and not interfere with subject measurement. A situation where one can reasonably expect and accept such interference, is one where changing the response alternatives would actually affect the meaning or content of the item (e.g. 'do you attend football ... rarely / sometimes / often' vs. 'do you attend football ... with little pleasure / ... / with much pleasure').

The problem of (non-)interference 'after the fact' concerns the question whether or not combining answers in different response categories would substantially affect subject measurement.

We give a formal definition of 'non-interference' later. The simplest way to study it is to investigate the effect of dichotomization. Most results are easily generalized to any (re-)partitioning of the response continuum.

1.3. LOSS OF INFORMATION VS. SUBSTANTIVE EFFECTS OF RESPONSE FORMAT

There has been some theoretical discussion as well as empirical research on the effects of the response format on subject measurement, notably with respect to the number of categories. There is also some evidence that the distribution of the responses is affected by the formulation of the response categories (cf. Berg & Rapaport (1954); Van Heerden & Hoogstraten (1979); quoted in Hoogstraten (1979), 60-62). Both the theoretical discussion and the empirical research were mostly concerned with statistical aspects, e.g. reliability and (loss of) information, and not so much with the question whether or not the number and phrasing of response alternatives interferes with subject measurement per se. By this we mean whether or not one is measuring the same construct when different kinds and numbers of graded response alternatives are used with the same items. For the present discussion we assume and accept that the phrasing and number of graded response alternatives in questionnaires will affect the precision and reliability of measurement, which can, in principle, be compensated for by the number of items. If, however, the phrasing and number of graded response alternatives affects what is being measured, we are confronted with a substantive problem of item response theory.

We take the position that the phrasing and number of graded response categories should not interfere with what is being measured. We take this position as a methodological working hypothesis. Effects of response category structure on precision, discriminability, reliability and related properties of measurement devices have no substantive importance.

1.4. THE MODEL APPROACH IN THE STUDY OF INTERFERENCE

Since we are concerned with the effect of polycho- or dichotomization on subject measurement, we must at least start the analysis on the basis of some item response model by which subject parameters can be estimated, and which can be tested empirically. Statements about subjects' latent trait scores would be gratuitous without a testable item response model. Starting from such a model, two approaches are open. One approach is strictly empirical: we can analyse data before and after dichotomization and see how subject measurement is affected. This approach is questionable on logical grounds: estimation of subject parameters is always conditional upon the assumed validity of the model, and if a model would not be compatible with dichotomization in the first place, the result of its application to dichotomized data would be meaningless. This leads us to adopting the model approach. By this we mean that we investigate the predictions of a model if data were dichotomized, and see under which conditions, if any, subject measurement is invariant under different polycho- or dichotomizations.

The joining assumption. The probability of subjects' responding with category j or category k is equal to the probability of responding with category h if category h replaces the categories j and k. In other words: if in one response format, the subject can choose between e.g. 'disagree', 'agree' and 'strongly agree', and in another response format he can choose between 'disagree', and 'agree', the probability of responding 'agree' in the latter case will equal the sum of the probabilities of responding 'agree' or 'strongly agree' in the former case. Mild and obvious though this assumption may seem, it has the strong implication that there is no difference between dichotomization before the fact and dichotomization after the fact. This joining assumption is of course trivially true if responses are dichotomized after the fact.

The principle of non-interference. The principle of non-interference can be formulated in terms of whether the subject parameter, ξ, in some measurement model, M, with additional parameters, θ, referring to items and/or categories, is affected by the partitioning of the response continuum. Let P and P' be two partitionings of the responses to a set of items. Non-interference can be defined in a formal way as follows.

> If P satisfies $M(\xi,\theta)$, and P' satisfies $M'(\xi',\theta')$, non-interference is defined by $\xi = t(\xi')$, in which t is an admissable transformation of the scale of ξ.

This definition does not explicitly require that M and M' are the same model. However, if we consider the item response model as a formalized theory of the data generating process, it would run counter to the very principle of theorizing, if we were forced to assume that different response models (apart from mere changes in the item- and category-parameters) would apply to any particular partitioning of the response continuum. In that case, the structure of the models M and M' would depend in an ad hoc way on the particular response format. So, if M is the unidimensional polychotomous Rasch model holding for P, and P' is a dichotomization, we expect the dichotomous Rasch model to hold for P'.

Formal definition of dichotomization. Suppose we have a polychotomous response model of the form:

$$P(g|v,i) = f(v,i,g) \tag{1}$$

where $v = 1,\dots,N$ refers to subjects,
$i = 1,\dots,k$ refers to items, and
$g = 1,\dots,m$ refers to response categories

and $P(g|v,i)$ denotes the probability that subject v chooses category g at item i.

By a dichotomization $D[h]$ is meant a joining of response alternatives such that alternatives $g=1,\dots,h-1$, are coded as 'no' or 'negative', and alternatives $g=h,\dots,m$, are coded as 'yes' or 'positive'.

Under the Joining Assumption, non-interference is satisfied if and only if

$$P(+|v,i,D[h]) = f^{*}(\xi'_v,i,h) = \Sigma\, f(\xi_v,i,g) \qquad (g=h,\dots,m) \tag{2}$$

with ξ'_v being an admissable transformation of ξ_v.

Furthermore, we use the following notation:

$$A = \{a_{vih}\},$$

$$a_{vih} \begin{cases} = 1, \text{ if subject v responds with category h at item i} \\ = 0, \text{ otherwise,} \end{cases}$$

$$A[h] = \{a^{D[h]}_{vi}\}, \tag{3}$$

$$a^{D[h]}_{vi} \begin{cases} = 1, \text{ if subject v responds with either category } g=h,\dots,m \text{ at item i,} \\ = 0, \text{ otherwise.} \end{cases}$$

2. THE POLYCHOTOMOUS RASCH MODEL (PRM)

2.1. THE PRINCIPLE OF SPECIFIC OBJECTIVITY

The Rasch model is based on the principle of specific objectivity in comparisons. It appears difficult to give an exact definition of this principle. It is concerned with comparative quantitative statements about properties of objects. Here, we present it first in ordinal terms: in determining whether e.g. one subject's attitude towards socialism is more positive than another subject's, we ask for their positive or negative responses towards a sample of issues from the pertinent universe of content. The comparative judgment is said to be specifically objective when it is independent of the particular sample of issues. In terms of probabilistic item response theory, specific objectivity means that the estimated difference (or ratio, depending on the scale type) of subjects' latent trait scores is independent of the (unknown) item parameters. In this formulation, specific objectivity is equivalent to sample independence. It should be noted that sample independence is not merely an empirical matter, as it seems to be considered by Lord((1980), 34 seqq.), but a theoretical property of the item response structure, and as such expressed in the form of the item response model. Whether or not data correspond to the model is of course a matter of empirical test.

In probabilistic item response theory, sample independence or specific objectivity means that subjects' parameters can be estimated independent of the item parameters, and vice versa, and this in turn means that there exist sufficient statistics for the item parameters which are independent from the subject parameters, and vice versa. From this requirement follows that the item response model must belong to the exponential family with separable parameters (cf. Fischer (1974), 415-420). Roskam (1983) and Jansen (1983) have given a different derivation of the unidimensional dichotomous Rasch model, not based on statistical considerations (see Roskam & Jansen (1983)).

2.2. THE MULTIDIMENSIONAL AND THE UNIDIMENSIONAL PRM (MPRM and UPRM)

Rasch (1961) has derived the following specifically objective item response model for polychotomous data. It is called the multidimensional polychotomous Rasch model (MPRM), since each item and each subject is characterized by a parameter vector.

$$P(h|v,i) = \frac{\exp\{\xi_v^{(h)} - \sigma_i^{(h)}\}}{\sum_g \exp\{\xi_v^{(g)} - \sigma_i^{(g)}\}} \tag{4}$$

where $v = 1,\ldots, N$ refers to subjects,
$i = 1,\ldots, k$ refers to items, and
$g,h = 1,\ldots, m$ refers to response categories.

It should be noted that the model assumes nominally the same response categories with all items. Each response category reflects a latent trait of its own. Therefore, we feel that this model should rather be called a multitrait model than a multidimensional model.

The MPRM is estimable with the following identifying restrictions (Fischer (1974), 432):

$$\begin{aligned} \xi_v^{(m)} &= \sigma_i^{(m)} = 0 \qquad \text{for all } v \text{ and } i \\ \sum_i \sigma_i^{(h)} &= 0 \qquad \text{for all } h = 1,\ldots,m \end{aligned} \tag{5}$$

The sufficient statistics for the subject parameters, $\xi_v^{(h)}$, in the MPRM are:

$$\sum_i a_{vih} = A_{vh} \tag{6}$$

where A_{vh} is the frequency of subject v responding with category h.

If the $\xi_v^{(h)}$ and $\sigma_i^{(h)}$ are linearly dependent, the following unidimensional reduction is possible (Rasch (1961), 330):

$$\begin{aligned} \xi_v^{(h)} &= \phi^{(h)}\xi_v + \psi^{(h)} \\ \sigma_i^{(h)} &= \phi^{(h)}\sigma_i \end{aligned} \tag{7}$$

leading to the unidimensional polychotomous Rasch model (UPRM):

$$P(h|v,i) = \frac{\exp\{\psi^{(h)}+\phi^{(h)}(\xi_v-\sigma_i)\}}{\sum_g \exp\{\psi^{(g)}+\phi^{(g)}(\xi_v-\sigma_i)\}} \tag{8}$$

with the following identifying restrictions (Fischer (1974), 439):

$$\begin{aligned} \phi^{(1)} &= 0 \\ \phi^{(m)} &= 1 \\ \psi^{(m)} &= 0 \\ \sum \sigma_i &= 0 \end{aligned} \tag{9}$$

In the unidimensional as well as in the multidimensional model, the response categories are the same with all items.

It should be noted that the dichotomous Rasch model is merely a special case of the UPRM. If m=2, eq. (9) and (8) reduce to the familiar dichotomous Rasch model.

The sufficient statistic for the subject parameter in the UPRM is:

$$\sum_h \phi^{(h)} A_{vh} \tag{10}$$

Therefore, the ϕ parameters are also called scoring parameters (Rasch (1961) 329). The ψ parameters can be considered location parameters. The meaning of the parameters is clarified by Fischer((1974), 437),as follows.

The preference of subject v for the response expressed by category h is a function of the subject parameter, and two other parameters which characterize the response category. The parameter $\phi^{(h)}$ weights the subject parameter differently for different categories. The parameter $\psi^{(h)}$ increases or decreases the subject's tendency to respond by category h. So, it expresses the 'easiness' or the appeal of that category in itself.

A special case arises when the scoring parameters are of the form $\phi^{(h)}=h-1$, meaning equidistant scoring.

Andrich's (1978) Rating Rasch Model is a special case of the UPRM. Both contain equidistant scoring as a sub-case. Whether or not data satisfy the equidistant scoring condition is an empirical matter, but Andersen (1977) has shown that equidistant scoring parameters can be derived as a methodological requirement, if we want a sufficient statistic for the subject parameter which is also quasi-continuous.

The category characteristic response curves (CCC's) of the UPRM, for the special case of equidistant scoring parameters, are shown in Figure 1.

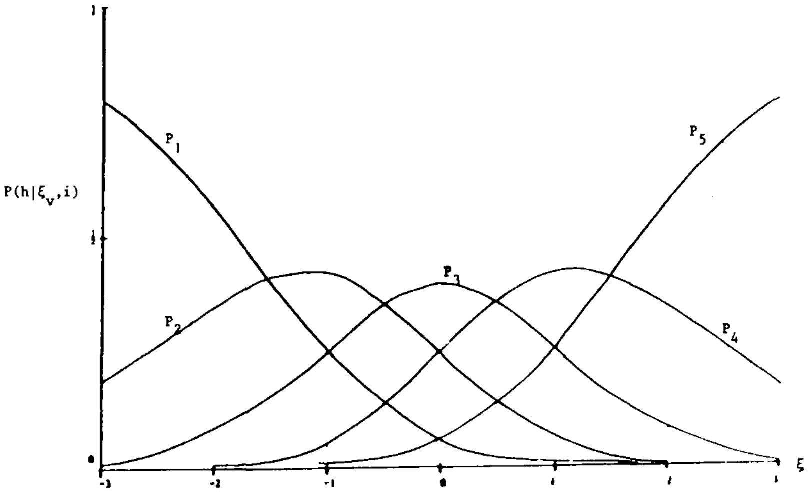


Figure 1

Category characteristic curves of an UPRM with ψ's equal to 0, 1.5, 2, 1.5 and 0 and with ϕ's equal to 0, 1, 2, 3 and 4 (equidistant scoring); σ_i is equal to 0.

2.3. THE UPRM AND DICHOTOMIZATION OF A GIVEN DATA

Suppose some multicategory data set, A, satisfies the UPRM. Let A[h] be a dichotomization of A, where a_{vih} and A[h] are defined by eq. (3).

Theorem: If A satisfies the UPRM for some $\phi^{(g)} \geq \phi^{(g-1)}$, i.e. for weakly ordered $\phi^{(g)}$, g=1,...,h,...,m, then the dichotomized data set A[h] will satisfy a dichotomous Rasch model if and only if

$$\phi^{(g)} \begin{cases} = a \text{ for } 1 \leq g \leq h-1 \\ = b \text{ for } h \leq g \leq m \end{cases} \qquad (11)$$

(where a=0 and b=1 as identifying constraints),

and the dichotomized data will satisfy:

$$P(+|v,i,D[h]) = \frac{\exp\{\xi_v-\sigma_i-\eta_{h-1}\}}{1+\exp\{\xi_v-\sigma_i-\eta_{h-1}\}} \tag{12}$$

where

$$\exp\{-\eta_{h-1}\} = \frac{\sum\limits_{g=h}^{g=m} \exp\{\psi^{(g)}\}}{\sum\limits_{g=1}^{g=h-1} \exp\{\psi^{(g)}\}} \tag{13}$$

Proof.

We assume polychotomous data which satisfy the UPRM. After dichotomization, these data will satisfy the dichotomous Rasch model, if and only if (cf. eq. 2):

$$P(+|v,i,D[h]) = \sum_g P(g|v,i), \quad g=h,\ldots,m \tag{14}$$

where $P(g|v,i)$ is as in eq. (8), and $P(+|v,i,D[h])$ is of the same form as eq. (8) with $m=2$, except for an admissable scale transformation.

Sufficiency. The proof of sufficiency is simple: substituting (11) in (8) leads immediately to (12) and (13).

Necessity. The proof of necessity is more complicated. In order to show that (11) is necessary it suffices to show that it is necessary if $\xi_v=\sigma_i$ in (8). In that case, eq. (14) immediately leads to (13); so the latter is necessary. Obviously, (11) is necessary for $m=2$. Next assume that a new category is added to the 'no'-set or the 'yes'-set of a dichotomized data set, and assume that (14) holds. It can then be shown that the scoring parameter of the added category is either equal to a, or equal to b (the scoring parameter of the 'no'-category, or of the 'yes'-category, resp.). For details of the proof, the reader is referred to Jansen (1983).

The category characteristic curves for the case of a dichotomizable UPRM are shown in Figure 2.

If the condition of the theorem is met, the parameter η_{h-1} is not identifiable. The dichotomization effects a shift of the latent scale of subjects and items. This shift is illustrated in Figure 3, for two hypothetical dichotomizations.

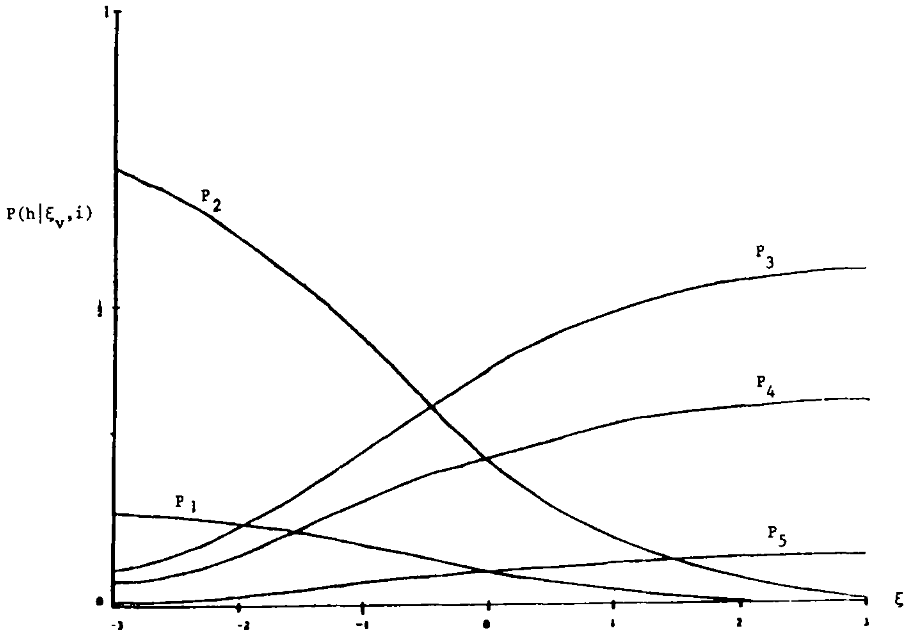


Figure 2

Category characteristic curves of an UPRM with ψ's equal to 0, 1.5, 2, 1.5 and 0, and ϕ's equal to 0, 0, 1, 1, 1. σ_i is equal to 0.

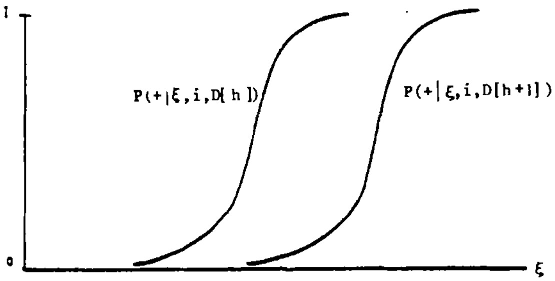


Figure 3

Response curves for item i in dichotomizations D[h], and D[h+1]

A closely related theorem can be formulated with respect to the joining of any two categories. If data satisfy a UPRM for m categories, and two (semantically adjacent, or not) categories, say j and k, are joined 'after the fact' to make a new category, say h, and if a UPRM for m-1 categories and with the same subject and item parameters is to hold for the data after joining j and k, then the parameters should satisfy:

$$\lambda_{ivh} = \lambda_{ivj} + \lambda_{ivk} \tag{15}$$

where

$$\lambda_{ivq} = \exp\{\psi^{(q)} + \phi^{(q)}(\xi_v - \sigma_i)\} \qquad q=h,j,\ k \tag{16}$$

We will show that eq. (15) can only hold with $\phi^{(h)}$ independent of ξ_v and σ_i if $\phi^{(j)} = \phi^{(k)}$.

Suppose eq. (15) holds, then, taking $\xi_v = \sigma_i$, it follows that:

$$\exp\ \psi^{(h)} = \exp\ \psi^{(j)} + \exp\ \psi^{(k)} \tag{17}$$

Using eq. (17), we find that eq. (15) implies:

$$\begin{aligned}(\exp\ \psi^{(j)} + \exp\ \psi^{(k)})\ \exp\{\phi^{(h)}(\xi_v - \sigma_i)\} &= \\ \exp\ \psi^{(j)}(\exp(\phi^{(j)}(\xi_v-\sigma_i))) &+ \exp\ \psi^{(k)}(\exp(\phi^{(k)}(\xi_v-\sigma_i)))\end{aligned} \tag{18}$$

The latter might be true if $\phi^{(h)}$ is some function of $\psi^{(j)}$, $\psi^{(k)}$, $\phi^{(j)}$ and $\phi^{(k)}$. Specifically, $\phi^{(h)}$ should be independent of ξ_v and σ_i, regardless of the values of $\psi^{(j)}$ and $\psi^{(k)}$.

Now let $\xi_v - \sigma_i = x$, $\phi^{(j)} = a$, $\phi^{(k)} = b$, and $\psi^{(j)}/(\psi^{(j)}+\psi^{(k)}) = p$, and $q=1-p$ for short. Then eq. (18) can be written as:

$$\phi^{(h)} = \frac{\ln(px^a+qx^b)}{\ln x} \tag{19}$$

The right member of (19) should be independent of x, in particular if $p=q=.5$. Differentiating with respect to x, and putting to zero, we find that (15) holds if and only if $\phi^{(j)} = \phi^{(k)}$. And if this condition is met, $\phi^{(h)} = \phi^{(j)} = \phi^{(k)}$.

Consequently, joining any two categories 'after the fact' will satisfy the UPRM with one category less and with the same subject and item parameters, if and only if the scoring parameters of the joined categories are equal. Repeatedly joining categories, we find that dichotomized data satisfy the Rasch model with the same subject and item parameters if and only if the scoring parameters take on only two different values.

Obviously, condition (11) imposes strong restrictions on the possibility of dichotomizing responses which satisfy the UPRM. Dichotomization does not interfere with subject measurement if and only if the scoring parameters take on only two different values, and if so, there is only one dichotomization which does not interfere with subject measurement. In fact, condition (11) means that the response categories belong to two groups, within each of which there is no discrimination, i.e. the subject's preference for one category over another is independent of the subject, but solely determined by the attraction of these categories in themselves.

2.4. THE IMPLICATIONS OF THE POLYCHOTOMOUS RASCH MODEL

As can be seen from the formulation of the polychotomous Rasch model, it is formally analogous to Bradley - Terry - Luce's choice model for the choice of one alternative from a set of response alternatives. It implies that the probability of a certain choice is a function of the relative strengths or utilities of each alternative. When alternatives are combined, e.g. by dichotomizing the set of alternatives, the utility of a combined subset of alternatives would be equal to the sum of the utilities of alternatives in the subset. Let h refer to the union of two alternatives, j and k. Equation (15) expresses our intuitive interpretation of repartitioning the response set, and it justifies its dichotomization: when we collaps two or more response alternatives into a single one, we take it that the probability of choosing the combined one is equal to the sum of the probabilities of choosing any of the alternatives entering into the combination. However, the resulting λ_{ivh} is no longer decomposable in the same way as the original λ_{iv}, unless the scoring parameters are equal.

A different problem concerns the effect of replacing a set of response options by a different set of response options, for instance by omitting one response option, e.g. 'don't know'. At first glance, it would seem reasonable to assume that this would not affect the parameters determining the response probabilities of the remaining options; it would mean that response probabilities are redistributed over the options in accordance with the constant ratio rule implicit in the BTL structure of the polychotomous Rasch model. This rule would be violated, however, if the subject would react to the omission of an option by choosing an option next to it.

But adding or deleting response alternatives in the response format of common attitude questionnaires does not imply adding or deleting possible responses, but rather means that segments in the response continuum are split up or are combined.

An obvious, but easily ignored property of the UPRM is that it does not in any way, imply an a priori ordering of the response alternatives. While this ignoring response order is only natural in the multidimensional model, the unidimensional model, in this respect, is not different from the multidimensional model. Due to the unidimensional reduction in eq. (7), the ordering of response alternatives is established through the analysis of the data. In view of this, the model cannot take into account any a priori established properties of the response alternatives, and would therefore not necessarily be compatible with joining categories on the basis of some a priori property.

In our assumption of joining response categories we have assumed that repartitioning or dichotomization 'before the fact' is not different from repartitioning or dichotomization after the fact. The Joining Assumption is trivially satisfied if given data are dichotomized. We found that the UPRM is not compatible with it. It may, however, be that the UPRM is compatible with different partitionings of the response continuum, if they are presented 'before the fact',and the subject is free to distribute his responses differently when a different response format is offered to him. Obviously, when response options are deleted or new ones added, the UPRM predicts that the probabilities of responses falling into categories which are common to two response formats, satisfy the constant ratio rule, unless a partial change in the response format would affect the parameters of all response categories.

2.5. CASES WHERE UPRM MIGHT APPLY

From the foregoing, we conclude that the UPRM may be valid in those cases where response alternatives are not graded, but are different qualities of response, reducible to a unidimensional latent trait. This interpretation is directly in line with the derivation of the UPRM from the MPRM. By intuitive reasoning, we would believe that the UPRM only applies to response alternatives which do not semantically or logically exclude one another, or, in other words, where the item cum response categories might be seen as the combination of several yes/no questions into a forced choice format. The following might be an example:

> If you were asked unexpectedly to speech on a social occasion, would you ...
>
> try to evade the appeal
> feel nervous
> feel flattered
> feel challenged .

Each of the possible responses can be a binary item in itself, which is logically independent from the others. Whether or not a unidimensional trait (e.g. 'poise' or 'assertiveness') underlies the probability of a positive response is an empirical matter. Besides, one would not feel justified in combining response categories because of their obvious qualitative differences. In this example, it can also be expected that omitting one alternative from the multiple choice format would lead to a redistribution of the response probabilities according to the constant ratio rule.

3. CONCLUSIONS AND CONSEQUENCES

We have shown that data satisfying the unidimensional polychotomous Rasch model with a certain number of categories, will not do so when responses categories are joined. In other words: the composition of the response format affects what is being measured in a non-trivial manner. In so far the composition of the response format is arbitrary,this implies that UPRM cannot be a methodological sound item response model. The same is true for models which are special cases of the UPRM, such as Andrich's (1978) Rating Rasch Model, and Master's (1982) Partial Credit Model. The only polychotomous model we have found in the literature which appears to be invariant under joining of graded response categories, is Samejima's (1969) Logistic Difference Model. Furthermore, the well-known equidistant Likert-scoring of graded responses suffers from a similar drawback. These and related issues will be more fully discussed in a forthcoming paper.

REFERENCES

[1] Andersen, E.B., Conditional inference for multiple choice questionnaires, British Journal of Mathematical and Statistical Psychology, 26 (1973), 31-57.

[2] Andersen, E.B., Sufficient statistics and latent trait models, Psychometrika, 42 (1977), 69-81.

[3] Andrich, D., A rating formulation for ordered response categories, Psychometrika, 43 (1978), 561-573.

[4] Berg, I.A. & Rapaport, G.M., Response bias in an unstructured questionnaire, The Journal of Psychology, 38 (1954), 475-481.

[5] Coombs, Cl.H., Theory of data, New York, Wiley (1964), reprinted, Ann Arbor: Mathesis Press (1976).

[6] Fischer, G.H., Einführung in die Theorie psychologischer tests, Bern Huber (1974).

[7] Guttman, L., The development of nonmetric space analysis; a letter to John Ross, Multivariate Behavioral Research, 2 (1967), 71-82, Reprinted in J.C. Lingoes, Edw. E. Roskam, & I. Borg (Eds.), Geometric representations of Relational Data; Readings in Multidimensional Scaling, Ann Arbor: Mathesis Press, 237-248.

[8] Heerden, J. Van & Hoogstraten, J., Response tendency in a questionnaire without questions, Applied Psychological Measurement, 3 (1979).

[9] Hoogstraten, J., De machteloze onderzoeker, Meppel, Boom (1979).

[10] Jansen, P.G.W., Rasch analysis of attitudinal data, Doct. diss. University of Nijmegen/Rijks Psychologische Dienst, Den Haag (1983).

[11] Likert, R., A technique for the measurement of attitudes, Archives of Psychology, 140 (1932).

[12] Lingoes, J.C., The rational of the Guttman-Lingoes nonmetric series: a letter to Doctor Philip Runkel, Multivariate Behavioral Research, 3 (1968), 495-508, Reprinted in J.C. Lingoes, Edw. E. Roskam, & I. Borg (Eds.), Geometric Representations of Relational Data; Readings in Multidimensional Scaling, Ann Arbor: Mathesis Press, 249-261.

[13] Lord, F.M., Applications of item response theory to practical testing problems, Hillsdale, Lawrence Erlbaum (1980).

[14] Masters, G.N., A Rasch model for partial credit scoring, Psychometrika, 47 (1982), 149-174.

[15] Rasch, G., On general laws and the meaning of measurement in psychology, Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, June 20 - July 30, 1960, Berkeley and Los Angeles, University of California Press (1961), 321-333.

[16] Roskam, Edw.E., Allgemeine Datentheorie, In H. Feger & H. Bredenkamp, Messen und Testen, Band 3 der serie Forschungsmethoden der Psychologie der Enzyklopädie der Psychologie, Göttingen, Hogrefe, 1-135.

This Page Intentionally Left Blank

This Page Intentionally Left Blank

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

COMPONENT ANALYSIS OF TRANSITION PROBABILITIES AND ITS APPLICATION TO PRISONER'S DILEMMA

Kullervo Rainio
University of Helsinki

The matrix of transition probabilities is examined as a weighted sum of component matrices. A least square method for determining the weights is shown.

The method of component analysis is applied to the results of Prisoner's Dilemma game. The player's expectations for the other player's strategy choices are analyzed by using 5 motivational components: 1) random expectation, 2) expecting competitive behavior, 3) repeating the expectations, 4) expecting "tit-for-tat" behavior, and 5) expecting repeated behavior. The components explain 55-82% of the variation of expectations. - The player's strategy choices are analyzed by using 5 motivational components: 1) random choice, 2) competitive choice, 3) repeating the choice, 4) choosing by the "tit-for-tat" method, and 5) choosing in an 'equalitarian" way according to the expectations. The components explain 54-79% of the variation of the strategy choices.

The weights of components of individual transition probabilities are also determined and their reliabilities computed. The highest reliabilities are over .80. The means and standard deviations of the individual weights of components are also given.

The general suitability of component analysis is briefly discussed. - A list of computer programs available is also given.

1. COMPONENT ANALYSIS OF TRANSITION PROBABILITIES

1.1. AN INTRODUCTORY EXAMPLE

Let us assume the following transition probabilities from states s_1, s_2, s_3 and s_4 at time point t to the states x_1 and x_2 at time point t+1:

$$\begin{array}{cccc} & & \multicolumn{2}{c}{t+1} \\ & & x_1 & x_2 \\ t & s_1 & .56 & .44 \\ & s_2 & .49 & .51 \\ & s_3 & .26 & .74 \\ & s_4 & .17 & .83 \end{array}$$

It is possible to think of this matrix as the weighted sum of several matrices, called components below. The four component matrices to be used in this example are the following:

$$\begin{array}{ccccccccc} & \multicolumn{2}{c}{\text{comp.1}} & \multicolumn{2}{c}{\text{comp.2}} & \multicolumn{2}{c}{\text{comp.3}} & \multicolumn{2}{c}{\text{comp.4}} \\ & x_1 & x_2 & x_1 & x_2 & x_1 & x_2 & x_1 & x_2 \\ s_1 & .5 & .5 & 0 & 1 & 1 & 0 & 1 & 0 \\ s_2 & .5 & .5 & 0 & 1 & 1 & 0 & 0 & 1 \\ s_3 & .5 & .5 & 0 & 1 & 0 & 1 & 1 & 0 \\ s_4 & .5 & .5 & 0 & 1 & 0 & 1 & 0 & 1 \\ & \multicolumn{2}{c}{w_1=?} & \multicolumn{2}{c}{w_2=?} & \multicolumn{2}{c}{w_3=?} & \multicolumn{2}{c}{w_4=?} \end{array}$$

Our task is to determine the weights w_1-w_4 in such a way that the weighted sum of the component matrices (see above) shall be as good an estimate as possible for the transition probability matrix given in the beginning.

The components have been formally chosen according to the principle of orthogonality. Such orthogonal components also seem to be easy to interpret. If the states are interpreted as behavioral states between which an agent can choose, then the transition probabilities represent tendencies to choose between the alternatives (x_1 and x_2 in our example). The components given above can be interpreted:

1) Component 1 -tendency to choose at random between x_1 and x_2 regardless of state at time point t.
2) Component 2 -tendency to choose x_2 regardless of state at time point t.
3) Component 3 -tendency to choose x_1 when the agent is in state s_1 or s_2 at time point t but x_2 if she is in state s_3 or s_4.
4) Component 4 -tendency to choose x_1 when the agent is in state s_1 or s_3 but x_2 if she is in state s_2 or s_4.

The tendencies described above may become more meaningful if the states are operationally defined. Actually, the figures given above are empirical relative transition frequencies - results of a multi-play Prisoner's Dilemma game - described in more detail in the second part of this report.

In the Prisoner's Dilemma game reported here, the player was even asked to give his expectation at each trial.

We shall use the symbol A for the partner's cooperative choice in the game and B for his noncooperative choice. Correspondingly, A' means that the player expects the partner to choose A and B' that the player expects B to be his partner's choice.

In our example states have the following meaning:

s_1 = A'A (i.e., at time point t the player has answered that she expects the partner to choose A on the following trial at time point t+1 and the partner has chosen A at time point t)
s_2 = A'B
s_3 = B'A
s_4 = B'B.

The column states x_1 and x_2 have the meaning:

x_1 = A' at t+1 (i.e., the player's expectation at the following time point t+1 is A').
x_2 = B'.

Thus, the component 2 means the tendency to choose noncooperatively (competitively) in all situations. Component 3 is the tendency to repeat the same expectation. Component 4 corresponds to the tendency to expect the partner to choose in the same way as in the earlier trial.

If we were able to determine the weights to be used in summing the component matrices we could get information about the relative strength of those tendencies described above.

It is easy to solve the problem in our example. We have only to solve the following four linear equations:

$.5w_1 = .17$ (from the 4. row)
$w_2+.5w_1 = .44$ (from the 1. row)
$w_3+.5w_1 = .49$ (from the 2. row)
$w_3+.5w_1 = .26$ (from the 3. row)

The weights are:

$w_1=.34$ $w_2=.27$ $w_3=.32$ $w_4=.09$

The solution of equations was applicable to our example where we had just as many orthogonal components as we had rows in the matrix (i.e., 4). Usually we have less equations than rows. As a general solution, it is convenient to use a least square method $^{(1)}$ to determine the weights. In doing that, all the information the transition probability matrix contains is used.

1.2. LEAST SQUARE SOLUTION

For computations we form the following vectors and matrices:

E = vector of transition probabilities (corresponding to empirical relative transition frequencies). Only the last (right) column of the transition probability matrix is used. Dimensions of E are n and 1, n being 2, 4, 8, ... (m-th potency of 2, m being a positive integer).

C = combined component matrix where the columns are the last (right) columns of each component matrix. The component matrix 1 with elements .5 is not included. Thus, the dimensions of C are n and c-1 where c is the total number of components.

C' = the transpose of C.

W = the vector of weights.

In the beginning, the vector of transition probabilities is reduced by the minimum (m) of the elements of the transition probability matrix. (The minimum is the smallest element of E or 1-E). We shall indicate the reduced vector by R. Thus,

$$R = E - m$$

The weight of component 1 is:

$$w_1 = 2m$$

The other weights are determined matrix-algebraically in the following way:

---

(1) We mean by 'least square method' determining the estimate in such a way that the sum of squares of differences between the estimate and the empirical transition probabilities is minimum.

I wish to thank prof. Anders Ekholm, Dept. of Statistics, University of Helsinki, for his valuable help in the mathematical solution.

$W = (C'C)^{-1}C'P$

where C' is the transpose of C and -1 indicates the inverse of the matrix C'C.

W contains the weights of components from 2 to n. (The weight w_0 was determined in the beginning).

EXAMPLE

The least square solution of our example in 1.1 is shown in the following, step by step:

$$E = \begin{bmatrix} .44 \\ .51 \\ .74 \\ .83 \end{bmatrix}$$

Minimum = 1-.83=.17

$$R = \begin{bmatrix} .27 \\ .34 \\ .57 \\ .66 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$C' = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$C'C = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}$$

The inverse of C'C =

$$\begin{bmatrix} .75 & -.50 & -.50 \\ -.50 & 1 & 0 \\ -.50 & 0 & 1 \end{bmatrix}$$

$(C'C)^{-1}C'$ =

$$\begin{bmatrix} .75 & .25 & .25 & -.25 \\ -.50 & -.50 & .50 & .50 \\ -.50 & .50 & -.50 & .50 \end{bmatrix}$$

The weights $W =$

$$\begin{bmatrix} .34 \\ .27 \\ .31 \\ .08 \end{bmatrix}$$

(Compare to the solutions in 1.1)

If V = the weights from 2 to n then the estimate

$$F = .5 \cdot w_1 + CV = \begin{bmatrix} .43 \\ .51 \\ .74 \\ .82 \end{bmatrix}$$

1.3. THE USE OF COMPONENT ANALYSIS

In the second part of this report we shall apply the component analysis of transition probabilities to our Prisoner's Dilemma results. (The idea of this component analysis was actually born in dealing with those results). Generally, the results produced by experimental games seem to be suitable for component analysis.

Another field of psychological study where it is easy to form interpretable components and, thus, apply our analysis is the study of learning and forgetting in terms of Markov chains (see, e.g., Coombs-Dawes-Tversky, 291-306).

2. COMPONENT ANALYSIS OF PRISONER'S DILEMMA

2.1. EXPERIMENTAL SET-UP

Several studies on Prisoner's Dilemma have been made at the Department of Social Psychology of the University of Helsinki in the years 1981-83, the results of which have been analyzed by the component analysis reported here. All those tests have been made by micro-computers (Pet Commodore) but the set-up of the studies varies. Two of them correspond well to the 'natural setting', because the information of both player's choices was transmitted through a special network. These studies are abbreviated by PDN - i.e. PD-Net. In the other studies (MAR81, NOV81, and FEB83) the choices of the other player were simulated by the program. A detailed description of the simulation is given in an earlier report (Rainio (1981)).

Each pair played 40 games - except in the NOV81 study where only 10 games were played.

In the PDN-FEB83 and in the PD-FEB83 studies the pay-off matrix was shown in a regular form with pay-offs varying from game to game. In the other studies the matrix was shown in a decomposed form.

An essential feature in our experimental set-up (in all studies) was that before each choice of strategy the player had to answer the question: 'What do you expect the other player to choose?' After that she got the opportunity to choose her own strategy.

The subjects were students participating in the elementary course of social psychology at the university. They did not know the Prisoner's Dilemma game. In the 5 different studies altogether 136 subjects were tested.

The player's state on each game is described by three characteristics.

1) The expectation, A' or B', A' indicating the player expects the other to choose cooperatively and B' being the noncooperative (competitive) choice.

2) The choice of the strategy, X or Z, X indicating the cooperative and Z the noncooperative (competitive) choice.
3) The information about the simultaneous choice of strategy by the other player, A indicating a cooperative and B a noncooperative choice by the other.

Thus, there are 8 different states on each game numbered as follows:

0) A'XA - the player expects the other to choose A, chooses herself X, and is then informed that the other player choose A.

1) A'XB
2) A'ZA
3) A'ZB
4) B'XA
5) B'XB
6) B'ZA
7) B'ZB.

2.2. COMPONENTS. ANALYSIS ON THE GROUP LEVEL

We can easily apply our component analysis to such transition matrices which have 2 columns. It is natural to analyze

1) the expectations (columns A' and B') using all the 8 states (from A'XA to B'BZ) of the earlier game as rows of the transition matrix, and

2) the choices of the strategies (columns X and Z) using 8 states (from $X_tA_tA'_{t+1}$ to $Z_tB_tB'_{t+1}$) as row of the transition matrix.

Thus we have two 8×2 matrices to be analyzed, one for expectations, shown in Table 1 as 'transition probabilities, expectation, empirical', and another for the strategy choices (Table 1, 'choice, empirical').

Because we have computed the frequencies over the total group of subjects in a study, our component analysis will yield a mathematical model for the mean tendencies of individuals to form their expectations - as well as their choices, correspondingly - in each different state. (Later we shall analyze the individual tendencies).

The components we shall use in our analysis appear in matrix forms in Table 3. Note that they have different meaning for expectations and for choices. The psychological interpretations are also given in Table 3.

The results of the component analysis appear in Table 1 where the weights are shown ('weights, exp.' and 'weights, choice') as well as the estimates and the residuals. The chi-squares have also been computed. In those cases where the p-values show no statistically significant differences between the estimates and the empirical vectors there are not necessarily any more components to be taken into account. When the estimate differs significantly from the experimental matrix (e.g., PDN-FEB82, 'choices') the residuals may contain some additional source of variation of the transition frequencies and cannot be interpreted as random variation.

The explanatory power of the component model - calculated by an approximative method - assumes the values shown in Table 4. - According to our study the component model can explain about 70-85% of the deviation of the transition probabilities from the .5 values (i.e., from the values they should assume at random).

By comparing the weights of components we can draw some psychological conclusions.

1) The tendency to repeat own choice seems to be considerably high (choices, w_3=.25,.37,.18,.22,.25) but the tendency to expect the other player to repeat her choice rather low (expect., w_5=.01,.15,.20,.17,.05). Thus, the player behaves in a more persevering way than she assumes the other to behave.

2) The tendency to repeat one's own earlier expectation is considerably high, too (expect., w_3=.17,.35,.26,.13). This can be interpreted as a rather high

tendency to persevere, too.

3) The tendency to choose in a noncooperative (or competitive) way varies much (choices, w_2=.09,.13,.56,.24,.2), as well as the tendency to expect such behavior (expect., w_2=.28,.18,.20,.23,.01).

When analyzing the results on the group level we always face a problem in the interpretation of our mean results: are the subjects similar to each other, all behaving in the "mixed" way the model describes, or are the results simply means in an unhomogeneous group of individuals who actually behave in their own, differing ways.

To find answers to this kind of questions we have to analyze the results on the individual level - trying to find out the component weights characteristic to each individual.

2.4. COMPONENTS. ANALYSIS ON THE INDIVIDUAL LEVEL

There were usually 40 games for each subject in our study. Because there are, thus, 39 transitions per individual in the 8×2 cells of a transition table, the cell frequencies become quite small. In principle, we can apply our component analysis to the transition frequencies of one individual, too, but the weights of components we get, may vary considerably by chance.

To measure the reliabilities of the individual component weights, we need to divide our transitions into two series, odd and even transitions.

The correlation between the odd and even series for each component weight was computed and corrected for length according to Spearman-Brown formula. The reliability correlations yielded by this split-half method are shown in Table 5.

The reliabilities are mainly considerably high if we take into account that the weights are computed from rather small frequencies of transitions.

On the basis of our results we may conclude that component analysis opens a method for finding out measures in terms of component weights which characterize the behavior of a player in PD game quite reliably.

It is possible to see in both tables same characteristic change during the process.

1) The tendency to repeat own expectations (exp., w_3) increases slowly (Table 6). The tendency to expect the other to choose in a repeated way (exp., w_5) appears on the later phase (exp., w_5 is slightly negative in the

beginning as one can see from Table 7) and increases from block 1 to block 2 (see Table 8). - The tendency to repeat own choice (choices, w_3) increases slowly. All these tendencies seem to be symptoms of increasing fatigue of the subjects during the test.

2) Both the tendency to expect 'tit-for-tat' behavior (exp., w_4) and to choose according to this strategy (choices, w_5) are very strong in the beginning (see particularly Table 7). To be 'equalitarian' and to expect the other to behave in that way, too, seems to be, thus, the strongest (or most common) tendency in the very beginning. It seems to get extinguished very quickly in the course of the first block (w_4 of exp. decreases from .42 to .03 and w_5 of choices from .35 to .08). - It is interesting to notice that the extinction of the 'equalitarian tendency' does not lead to competitive behavior but simply to tiredness - the subject starts to repeat his expectations and choices.

3) Both the tendency to expect the other to behave competitively (exp., w_2) and to choose own strategy in that way (choices, w_2) remain quite constant during the test.

The component analysis does not lead to any very surprising results. Because it is able to give measures of those behavior tendencies in Prisoner's Dilemma already known, it makes the picture more clear.

3. DISCUSSION

Our study on Prisoner's Dilemma has to be seen merely as an example of the application of the component analysis introduced in this report, not as an attempt to analyze deeply the human motivation in that game. However, the component analysis may offer a good basis for making comparative analysis between results got from differing experimental games. In the future, it would be also interesting to find out how the component weights may change if some variation is systematically produced in subject's motivation in game situations by differing the instructions and the pay-offs.

The general applicability of component analysis was already briefly discussed in Chapter 1. Summarizing we may characterize the component analysis as a 'factor analysis of transition probabilities', because it gives a holistic picture of the factors 'behind' a set of transition probabilities - as factor analysis does in relation to separate correlation studies.

REFERENCES

[1] Coombs, C.H., Dawes, R.M., Tversky, A., Mathematical psychology, Prentice-Hall (1970).

[2] Kemeny, J.G., Snell, J.L., Thompson, G.L., Introduction to finite mathematics, Prentice-Hall (1957).

[3] Kuhlman, D.M., Wimberley, D.L., Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental games, Journal of Personality Soc. Psychol. 34 (1976), 69-81.

[4] Miller, D.T., Holmes, J.G., The role of situational restrictiveness on self-fulfilling prophecies: A theoretical and empirical extension of Kelley and Stahelski's triangle hypothesis, Journal of Personality Soc. Psyhcol. 31 (1975), 661-673.

[5] Misra, S., Triangle effect and the connotative meaning of trust in Prisoner's Dilemma: A cross cultural study, Internat. Journal of Psychol., 14 (1979), 255-263.

[6] Pruitt, D.G., Kimmel, M.J., Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future, Ann. Rev. Psychol., 28 (1977), 363-392.

[7] Rainio, K., Simulation of multi-play Prisoner's Dilemma according to a quantitative goal/expectation theory, Research Reports, Dep. of Soc. Psychol., University of Helsinki, 2 (1981).

[8] Rapoport, A., Chammah, A.M., Prisoner's Dilemma, University of Michigan Press (1965).

[9] Schulz, U., Hesse, W., A model for gaming behavior in sequences of Prisoner's Dilemma games in consideration of expectations, Psychologische Beitrage 20 (1978), 551-570.

```
TABLE 1.  Component analysis. PD

Program: COMPANPD-MAY83D DATA: PDNPPOB1BL-F82
Weights, exp.:     .27     .22     .27     .2      .06
Weights, choice:  .24     .11     .31     .13     .14

                       Transition probabilities:expectation

  Empirical                       Estimate                        Residuals

               A'     B'     Sum              A'     B'     Sum       A'      B'
  A'XA     .67    .33     40       A'XA    .64    .36     40       .03    -.03
  A'XB     .59    .41     64       A'XB    .58    .42     64       .02    -.02
  A'ZA     .39    .61     28       A'ZA    .44    .56     28      -.05     .05
  A'ZB     .38    .63     56       A'ZB    .38    .62     56      0       0
  B'XA     .37    .63     52       B'XA    .37    .63     52      -.01     .01
  B'XB     .27    .73     83       B'XB    .31    .69     83      -.04     .04
  B'ZA     .19    .81     77       B'ZA    .17    .83     77       .02    -.02
  B'ZB     .14    .86    248       B'ZB    .11    .89    248       .03    -.03

                          Transition probabilities:choice

  Empirical                       Estimate                        Residuals
           X      Z       Sum              X      Z       Sum      X       Z
  XAA'     .76    .24     46       XAA'    .77    .23     46      -.01     .01
  XAB'     .67    .33     45       XAB'    .63    .37     45       .04    -.04
  XBA'     .62    .38     60       XBA'    .64    .36     60      -.03     .03
  XBB'     .49    .51     86       XBB'    .5     .5      86      -.01     .01
  ZAA'     .38    .62     26       ZAA'    .46    .54     26      -.07     .07
  ZAB'     .35    .65     79       ZAB'    .32    .68     79       .04    -.04
  ZBA'     .44    .56     55       ZBA'    .33    .67     55       .11    -.11
  ZBB'     .12    .88    249       ZBB'    .19    .81    249      -.07     .07

Chi-Squares
Expectations: 3.64   DF = 3

Choices: 12.17 DF = 3.
```

TABLE 2. Component weights. A summary

Expectations

Study:	Components: C1	C2	C3	C4	C5
PDN-FEB82					
Block 1	.34	.28	.17	.22	-.01
Block 2	.18	.18	.35	.21	.15
PDN-FEB83	.24	.20	.26	-.01	.20
PD-MAR81	.45	.23	.13	.08	.17
PD-FEB83	.40	.01	.30	.16	.05
(PD-NOV81)	.25	.20	.20	.34	.04

Choices

Study:	Components: C1	C2	C3	C4	C5
PDN-FEB82					
Block 1	.26	.09	.25	.17	.19
Block 2	.20	.13	.37	.17	.08
PDN-FEB83	.08	.56	.18	.07	.02
PD-MAR81	.30	.24	.22	.08	.18
PD-FEB83	.06	.42	.25	.14	.20
(PD-NOV81)	.22	.20	.01	.20	.35

In the study PD-NOV81, there were only 10 games per subject.

TABLE 3. Components and their psychological interpretations

	Component									
	1		2		3		4		5	
Row	0	1	0	1	0	1	0	1	0	1
0	.5	.5	0	1	1	0	1	0	1	0
1	.5	.5	0	1	1	0	1	0	0	1
2	.5	.5	0	1	1	0	0	1	1	0
3	.5	.5	0	1	1	0	0	1	0	1
4	.5	.5	0	1	0	1	1	0	1	0
5	.5	.5	0	1	0	1	1	0	0	1
6	.5	.5	0	1	0	1	0	1	1	0
7	.5	.5	0	1	0	1	0	1	0	1

Analysis of expectations:

The rows are states A'XA...B'ZB.

The columns are:

0) Expectation A' in the following game.

1) Expectation B' in the following game.

The interpretation of components:

Comp. 1) tendency to choose the expectation at random.

Comp. 2) tendency to expect noncooperative (competitive) choice - regardless of the state.

Comp. 3) tendency to repeat one's earlier expectation.

Comp. 4) tendency to expect that the other chooses in accordance with one's earlier choice, i.e., in the same way as the player herself has just chosen (expecting a 'tit-for-tat' rule from the other's part).

Comp. 5) tendency to expect that the other simply continues her choice strategy.

Analysis of choices:

The rows are the following states:

0) XAA' (i.e., X and A on the game t and A' on the game t+1).

1) XAB', 2) XBA', 3) XBB', 4) ZAA', 5) ZAB', 6) ZBA', 7) ZBB'.

The columns are:

0) X on the game t+1

1) Z on the game t+1.

The interpretation of components:

Comp. 1) tendency to choose at random

Comp. 2) tendency to choose noncooperatively (competitively).

Comp. 3) tendency to repeat own choice.

Comp. 4) tendency to choose in an 'echoing' way, i.e., in the same way as the other has just chosen ('tit-for-tat').

Comp. 5) tendency to choose in an 'equalitarian' way in accordance to one's expectation (i.e., to choose X if expectation is A and Z if it is B).

TABLE 4. Explanatory power of components

Study:	Expect.	Choices
FEB82N		
Block 1	73%	72%
Block 2	82%	68%
FEB83N	55%	70%
MAR81	74%	57%
FEB83	73%	75%
NOV81	64%	54%

TABLE 5. Reliabilities of components

Set-up of the studies:

FEB82N and FEB83N - communication net
M81F82 and NOV81 - Partner simulated.
(Correlations are corrected for length).

Expectations:

	Study			
Component	FEB82N	FEB83N	M81F83	NOV81
Comp. 1	.66	.77	.66	.75
Comp. 2	.63	.84	.69	.47
Comp. 3	.71	.56	.61	.63
Comp. 4	.37	.44	.50	0
Comp. 5	.12	.66	.17	-.44

Choices:

Comp. 1	.69	.67	.50	.56
Comp. 2	.67	.71	.69	.28
Comp. 3	.54	.53	.72	.38
Comp. 4	.61	.12	.38	.44
Comp. 5	.82	.60	.77	.53

TABLE 6. Constancy of components. Blocks 1 and 2

Means of weight (studies FEB82N, FEB83N, and MAR81)

Block 1: first 20 transitions.

Block 2: last 20 transitions.

Expectations

	w_1	w_2	w_3	w_4	w_5
Block 1	.43	.20	.14	.13	.10
Block 2	.31	.26	.26	.03	.18

Choices

	w_1	w_2	w_3	w_4	w_5
Block 1	.18	.29	.19	.10	.23
Block 2	.27	.26	.30	.07	.08

TABLE 7. Constancy of components. 4 first transitions (Computed from data MAR81, NOV81, and FEB82N).

1. transition: from game 1 to 2; 2. transition: from 2 to 3;etc.

Expectations:

	w_1	w_2	w_3	w_4	w_5
1. transition	.21	.19	.16	.42	-.01
2. transition	0	.37	.23	.33	-.07
3. transition	0	.52	.11	.17	-.21
4. transition	0	.38	-.14	.32	.16

Choices:

	w_1	w_2	w_3	w_4	w_5
1. transition	0	.26	.20	.19	.35
2. transition	0	.38	.22	.04	.36
3. transition	0	.19	-.01	.48	.32
4. transition	0	.31	-.03	.34	.33

TRENDS IN MATHEMATICAL PSYCHOLOGY
E. Degreef and J. Van Buggenhaut (editors)
© Elsevier Science Publishers B.V. (North-Holland), 1984

AN APPLICATION OF MULTIDIMENSIONAL SCALING ON PREDICTION: THE RADEX-STRUCTURE

Frank Van Overwalle
Vrije Universiteit Brussel, Belgium

The method we propose is suitable for large sets of variables, which have to be grouped in smaller units, in order to predict a criterion. Contrary to the more conventional methods, such as factor analysis, that do not account for the criterion to be predicted while grouping the data, we propose an application of multidimensional scaling, in order to find a RADEX in the data. Since a RADEX consists of (1) an hierarchy of circumplexes which divide low vs. high predictors, and (2) weak continguity regions which group together homogene sets of predictors in each circumplex, it thus reconciles both grouping and predictive purposes at once.

1. INTRODUCTION

The aim in psychological research often is to predict a criterion - a specific behaviour or outcome - through a number of variables. One hopes that the prediction of the criterion through these variables is as high as possible. This objective is often called: predictive validity. A frequently used statistical method to fulfil this purpose is: multiple regression analysis. A parallel purpose, which is often added to the first method above, is to clarify how this prediction is possible. One tries to build up a structural model which explains more clearly how events actually take place, in other words, a model which explains the causal relations between the predictors and the criterion. For this purpose path-analysis or other linear structural models (e.g. Lisrel) are used. One can summarise both predictive methods by considering regression-analysis a first-level approach, while structural models also include factors on a deeper causal level.

When the predictors are well-defined distinct concepts, there will be little problem to apply the statistical methods mentioned. On the contrary, when a large number of different variables are gathered (which is often the case in psychological research of an explorative nature) it is not very clear how these data should be selected in order to apply regression-analysis or causal models. Since all variables are still in their original form, it often means that their number is too large, or it means that their intercorrelations are too high, which leads to unreliable regression-coefficients. The researcher is often forced, first of all, to reduce the large number of variables to fewer units. These units must have a psychological meaning and must be justified by statistical means. This activity in which smaller separate units are sought for is called: discriminant validity (Denison (1982)). For that purpose factor- or clusteranalysis is often used. But these classical methods share an important disadvantage. While structuring their data, they do not account for the criterion that has to be predicted. The grouping of the data is not only far from optimal for prediction, sometimes it even conflicts with it. In this text, this conflict between discriminant and predictive validity is further elaborated on.

It is, indeed, possible that poor and strong predictors fall together in one factor or cluster, such that the final accuracy of this factor in prediction becomes moderate (figure 1A). One has little control over the occurance of such events.

The problem becomes even more acute when several strong predictors are available. Since strong predictors correlate very highly with the criterion, their intercorrelations must be rather high, too. The danger that all these strong predictors will be grouped together in one factor or cluster becomes very realistic. What should have been a multiple prediction/explanation of the criterion, is now concealed in one construct (figure 1B).

Denison (1982) states the problem in a similar way: "... a dilemma which is at the root of much ofbehavioural and attitudinal research: if two measures are to be taken as operalisations of conceptually distinct constructs, they must be at least somewhat unrelated. However, if a close predictive relationship actually exists between these two constructs, the two measures must be highly related".

We will now present a method, which to a large extent can reconcile the conflict between discriminant and predictive validity.

2. MULTIDIMENSIONAL SCALING

Our approach, in which both divergent purposes are solved at once, is based on multidimensional scaling (MDS). This approach actually consists of two steps.

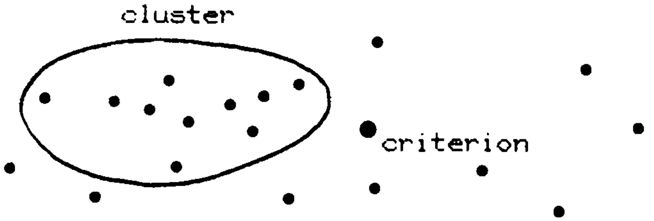


Figure 1A

Poor and strong predictors grouped together in one cluster/factor.

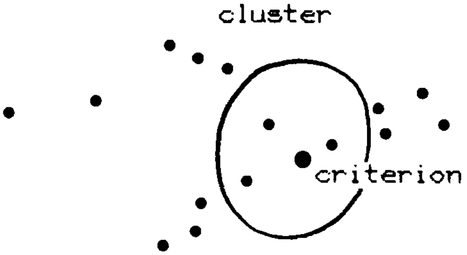


Figure 1B

Different strong predictors grouped together in one cluster/factor.

1) Monotone distance analysis (MDA), which enables us to look for a meaningful structure in the data. In this structure the criterion plays an essential role.

2) Confirmatory monotone distance analysis (CMDA) which enables to test the hypothetical structure against the original data.

In this text, we will first give a short introduction to the basic principles of both methods of MDS. Secondly, we will explain our specific approach used in prediction-research. And finally, we shall illustrate our approach by an application in the field of educational psychology.

2.1. MONOTONE DISTANCE ANALYSIS

Non-metric or monotone distance analysis (MDA) was first developed by Shepard (1962 ab) and further worked out by Kruskal (1964 ab). Later on, the Shepard-Kruskal approach initiated a vast research in other varieties of data-analysis, which are now grouped together under the heading of non-metric (or monotone) multidimensional scaling.

The basic purpose of MDA is to represent (dis)similarities (e.g. correlations) between variables in a spatial model. In this model, variables become points and (dis)similarities become distances. The closer two points, the higher the similarity between them. The more distant two points, the lower their similarity. A spatial representation renders the interpretation and analysis of complex structures between variables much easier. These structures can be factors, clusters or other forms of homogene groups, but other more complex structures are also possible. In our method we will apply such a more complex structure.

In our example we used MINISSA-I of Guttman, Lingoes and Roskam (Guttman (1968), Lingoes (1973), Lingoes and Roskam (1973)). Although apparent differences with the original method of Shepard, their method is also based on a monotone model.

Monotone means that the rankorder of the spatial distances should be identical to the rankorder of the original (dis)similarities. However, this is never really possible. The deviation from monotonicity is expressed by the alienation coefficient, which can take values between 0 and 1. The alienation can thus be considered as a measure of non-explained variance by the spatial model.

2.2. CONFIRMATORY MONOTONE DISTANCE ANALYSIS

While looking for a structure in the data, one should not always be led by the distances in the spatial representation. One can also start from a specific hypothesis about the relations between the data, and then try to find them back in the spatial representation of the data. Even if the chosen hypothesis is a very reasonable or tested one, it happens frequently that a minority of points doesn't fit the hypothetical structure. This is so because MDA is a "blind" method which only accounts for the (dis)similarities. Is it possible to find a spatial representation that is optimal with regard to the constraints of the hypothesis? And can this new representation be tested for its (dis)agreement with the original data? These two questions can be solved by a constrained/confirmatory monotone distance analysis (CMDA).

The CMDA-program that we used in our example was developed by Lingoes and Borg (Lingoes and Borg (1978), Borg and Lingoes (1979), (1980)). The program starts from the original spatial configuration obtained during the previous MDA. Then this original configuration is submitted to the constraints derived from the hypothesis. Both the constraints and the distances should be monotone invariant with regard to the original (dis)similarities. If the original configuration was not completely satisfying the hypothetical structure, the alienation of the newer, constrained configuration will increase. In other words, the constrained configuration will deviate even more from the original (dis)similarities.

Is this difference significant or not? Should we accept the chosen hypothesis or not? Lingoes and Borg developed two approaches to test the difference between both configurations, with and without constraints. The first statistical approach is an adaption of Hotelling's t-test (Lingoes and Borg (1980)). Later on, Lingoes and Borg (1982) have criticized the application of this test and in 1983, they developed an alternative "mathematical decision model" to test the equivalence between both configurations. In our example, we will describe these approaches in some more detail.

3. CONSTRAINTS IN PREDICTION-RESEARCH

In this part we will describe our particular application of MDS in prediction-research. This application will enable us to combine two purposes at once:

1) structuring of the variables in separate regions, e.g. discriminant validity;
2) selecting poor, moderate and strong predictors of the criterion, e.g. predictive validity.

We already mentioned that our approach consists of two phases:

1) looking for a structure or hypothesis after a "blind" MDA;
2) confirmation or rejection of the hypothesized structure by means of a CMDA.

The particularity of our approach lies in the structure we are looking for: the Radex (Guttman (1954)). A radex-structure will serve our purpose, because it can have both discriminant and predictive validity. We will now explain the characteristics of the radex in more detail, and show how these can be transformed into operational constraints for the CMDA-program.

3.1. THE RADEX-STRUCTURE

The MAD is based on a (dis)similarity-matrix between the variables. In our example, we used spearman-correlations, but other association-measures are also possible. Before entering the analysis, two preceding transformations are necessary:

1) variables with a negative correlation with the criterion are inversed.
2) based on the significance-level or similar criteria, variables with too low correlations with the criterion are excluded from the analysis. They should unnecessarily complicate the spatial representation. Variables above the selection are included. We shall now call them predictors.

A monotone distance analysis (MDA) is then performed on the correlation matrix of predictors and criterion. For reasons of visibility and interpretability of the spatial configuration, a plot of 2 dimensions is prefered. Because of the fixed selection-level of the predictors and the inversion of negative predictors, the criterion to predict will lie around the centre of the spatial plot.

In this spatial representation the radex-structure is sought:

1) the intensity of the correlation with the criterion can be seen from the distance to the criterion. To separate low, moderate, and strong predictors, one draws circles with the criterion as their common centre. These circles cut the spatial plot in circumplexes (Guttman (1954), see figure 2A).
2) to separate groups of highly intercorrelating and conceptually similar predictors, each circumplex is cut into regions, like a pie is cut into pieces (figure 2B).

A configuration with the above mentioned pie or wheel-form is a radex (Guttman (1954)). The parts of the radex are so-called regions (Lingoes (1981)).

Looking for a radex in a plot of 3 dimensions is of course much more difficult. So up till now we have prefered a higher alienation in two dimensions, instead of a lower alienation in three or more dimensions.

3.2. TEST OF HYPOTHETICAL RADEX-STRUCTURE

There are, of course, several ways for cutting up the space in a radex. Once a radex-structure found, it is just one of many possible radexes that can be hypothesized or can be hidden in the data. In many cases, not all the points will fit into the radex. With the CMDA-program, alternative hypotheses can be tested, and rejected or accepted.

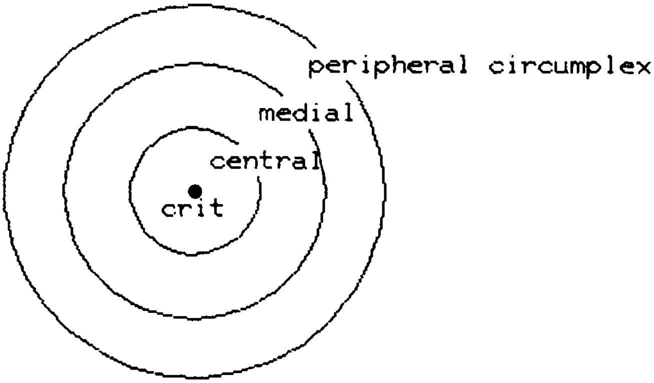


Figure 2A

Peripheral/medial/central circumplexes of poor/medium/strong predictors.

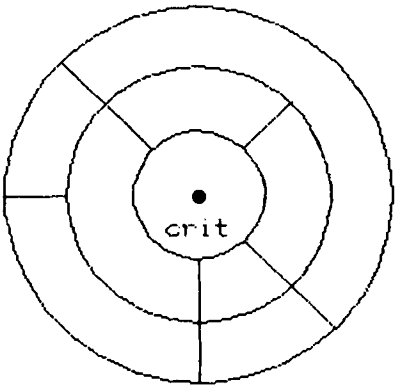


Figure 2B

Radex with different circumplexes, each of them cut in several regions.

Since CMDA is based on a non-metric model, the radex structure must be converted in ordinal regional constraints as input for the CMDA-program.

3.2.1. Constraints for circumplexes.

Points of central circumplexes are closer to the criterion than points of more periphere circumplexes (see figure 3A). The distances between points, i (in the circumplex, C_i) and the criterion, c, result in the following constraint:

$$\text{if}: C_i < C_j < \ldots < C_1 < c$$
$$\text{then}: d(i,c) > d(j,c) > \ldots > d(1,c)$$

3.2.2. Constraints for regions in circumplexes.

Lingoes (1981) developed a method to convert several forms of regional hypotheses in ordinal cosntraints. The keystones in his approach are the outer- and innerpoints of a region. These points are to be defined for each pair of regions, R and R^*.

A point of region R is an outerpoint, O, if it has the shortest distance with a point of region R^*. Going through all points of R^*, and for each of them taking the closest point of R (which then becomes an outerpoint), one can define all outerpoints of region R in regard to region R^* (see figure 3B). The outerpoints of R can be considered as constituting a boundary of R against the opposite region R^*. Consequently, the innerpoints of R are all points of R which are no outerpoints (see figure 3B).

In a similar way, all outerpoints and innerpoints of R^* can be defined. And this can be done for all pairs of regions in a circumplex. To prevent a situation where all points are outerpoints, one must at least include more than 2 points in each region.

Since regions in a circumplex are not always convex, they must comply with the constraints of a "weak continguity type A" (Lingoes (1981)). A region R satisfies this weak (type A) continguity constraint" if (*each*) innerpoint of R is closer to *some* outerpoints of R, than it is to any outerpoint of another region R^*" (Lingoes (1981), p290-291).

For our radex-structure, this definition of Lingoes is slightly adapted: ... if *each* innerpoint of R is closer to *one* outerpoint of R, than it is to any outerpoint of another region R^*.

One can check our slightly adapted definition in figure 3C by taking a target innerpoint, i^*, in region R and seeking the smallest distance with one outerpoint (the target outerpoint, o^*) of the same region R. This distance $d(i^*, o^*)$ must be smaller than the distances between this target innerpoint i^* of region R, and all other outerpoints of R^*. This is true for each innerpoint of R chosen as target.

To explain our arguments for the adaption of Lingoes' definition, we take figure 3C once again. As can be seen in this figure, only the distance $d(i^*, o^*)$ is always smaller than the distances between the target innerpoint i^* of R and all the outerpoints of R^*: o_3, o_4, o_5. The distance between the i^* of R and another outerpoint of R can be larger (e.g. $d(i^*, o_2) > d(i^*, o_3)$ in figure 3C). That's why only *one* outerpoint of R is actually chosen for comparison with the outerpoints of another region R^*. But in contrast with Lingoes (see example p.297 (1981), in a radex structure *each* innerpoint of R must satisfy the weak continguity constraints. To compute all the regional constraints of a radex outlined above, a REGION-program was written in Fortran 4 by Van Overwalle (1983).

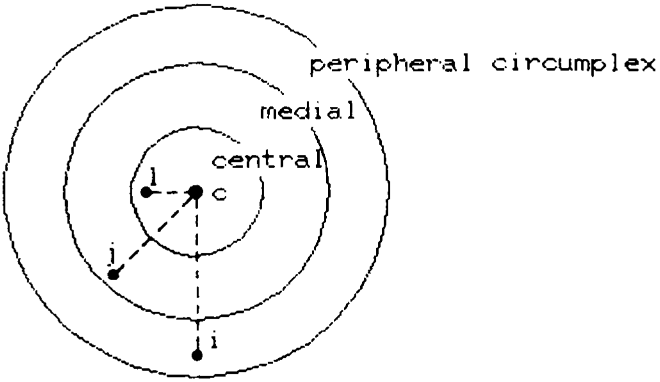

peripheral circumplex
medial
central
i
c
i
i

Figure 3A

Regional constraints for circumplexes (see text)

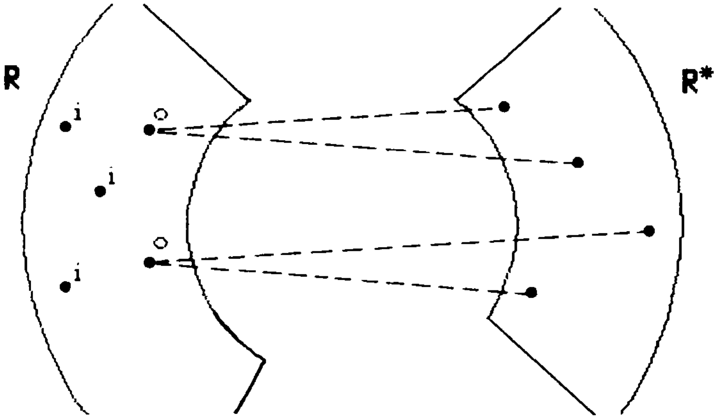


Figure 3B

Outerpoints and innerpoints of region R (see text)

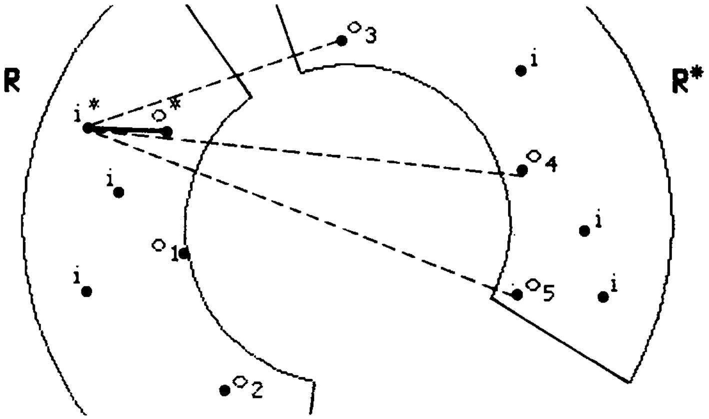


Figure 3C

Regional constraints for regions of weak contiguity type A (see text)

4. AN EXAMPLE

To illustrate our application of the radex-structure in prediction, we take an example from our research on student-achievement. This example includes data of 2 pilotstudies from firstyear students at university, gathered after their first half year at university.

We will now illustrate our approach step by step.

4.1. MONOTONE DISTANCE ANALYSIS

Only 46 variables were selected for an MDA-plot, by a significance-level of α = 05. The spatial configuration after a MINISSA is shown in figure 4. One can readily see that the final examination results, our criterion, lies in the centre of the plot. The most central circumplex is labelled "social contact". It is surrounded by two medial regions "self-esteem" and "motivation". Finally, one peripheral region "secondary school", is connected with the previous self-esteem region.

We have drawn our hypothetical structure by means of

- circles to separate low, medial and high circumplexes,
- straight lines to split up the "weak contiguity" regions.

Although our hypothetical radex generally seems to fit well in the plot, six points contradict our hypothesis. To be in line with our proposed radex, they must lie in another region, indicated by the arrows.

Since most violations occur in the centre of the MINISSA-plot, these violations can be a consequence of a reduced accuracy of a spatial representation, due to too much irrelevant points in the periphery of the plot. This argument seems to be confirmed by a second analysis where the inclusion level was increased to a correlation of 0.40 (see figure 5). Only 2 points in this second spatial plot violate the radex-structure, but our peripheral region of secondary school is now completely lost.

4.2. REGIONAL CONSTRAINTS FOR A RADEX

Another, more simple method to check a part of the hypothesized radex-structure, the circumplex, is shown in figure 6. We just listed all original correlations with the criterion, classified according to their hypothetical position in a region. Once again, there are several violations, but in general, the circumplex is confirmed. Moreover, we have now an exact idea of the boundaries between the circumplexes. The levels at which the circles in our plot are drawn, coincide with spearman-correlations of 0.40 and 0.70. The radex proposed in the first spatial configuration (figure 4), confirmed by a second spatial plot (figure 5) and by the original correlations (figure 6), is entered in the REGION-program (Van Overwalle (1983)) and results in 24 regional constraints. These constraints serve now as input for the Lingoes-Borg CMDA-program.

4.3. CONFIRMATORY MONOTONE DISTANCE ANALYSIS

Figure 7 represents a spatial plot of a confirmatory/constraints analysis. All points are now in their expected position according to the radex-structure outlined above, except for one point. It is not our intention to speculate on the causes of this exception in this text, since we present this analysis merely for illustrative purposes.

On the contrary, we will now discuss some approaches to evaluate the results of a confirmatory/constraints analysis, beyond the visual interpretability of the spatial configuration.

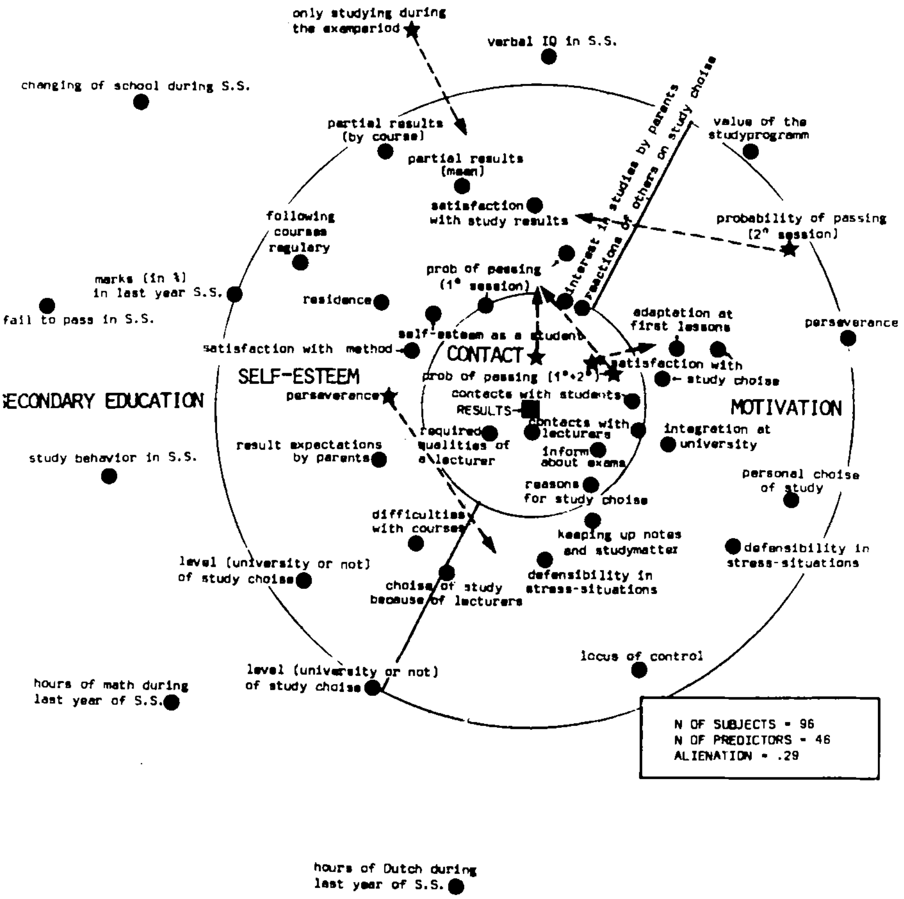


Figure 4

MINISSA

Spatial plot of spearman correlations ($\alpha \leq 0.5$) between predictors and criterion.

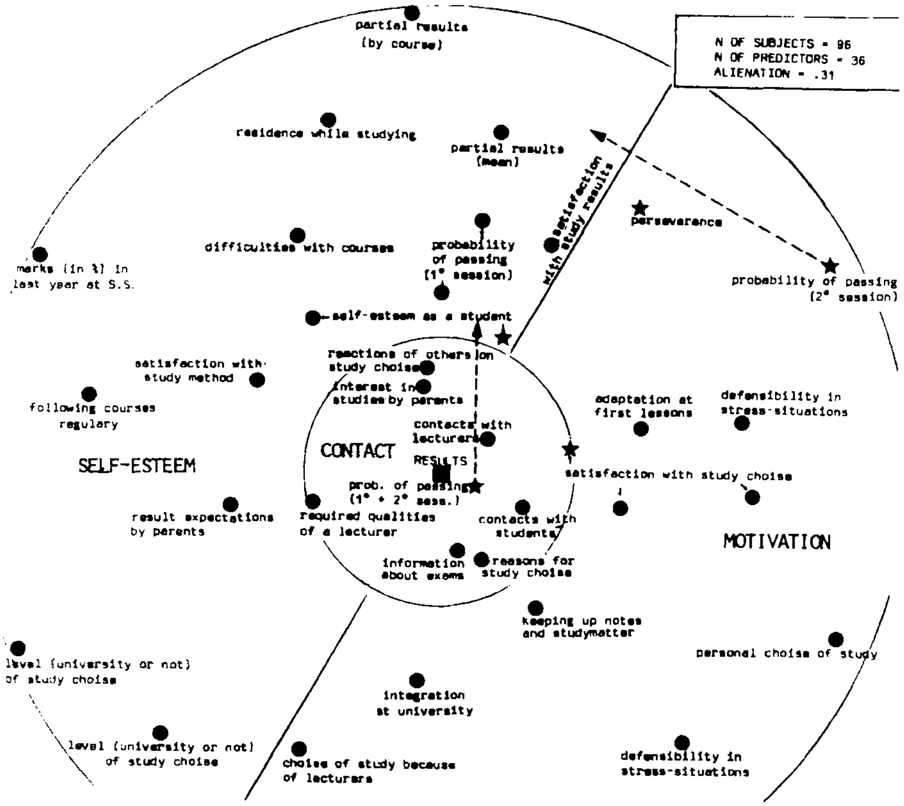

N OF SUBJECTS = 96
N OF PREDICTORS = 36
ALIENATION = .31
partial results (by course)
residence while studying
partial results (mean)
with satisfaction study results
perseverance
difficulties with courses
probability of passing (1° session)
marks (in %) in last year at S.S.
probability of passing (2° session)
self-esteem as a student
reactions of others on study choise
satisfaction with study method
interest in studies by parents
following courses regulary
adaptation at first lessons
defensibility in stress-situations
contacts with lecturers
CONTACT
RESULTS
SELF-ESTEEM
satisfaction with study choise
prob. of passing (1° + 2° sess.)
result expectations by parents
required qualities of a lecturer
contacts with students
MOTIVATION
information about exams
reasons for study choise
keeping up notes and studymatter
personal choise of study
level (university or not) of study choise
integration at university
level (university or not) of study choise
choise of study because of lecturers
defensibility in stress-situations

Figure 5

MINISSA

Spatial plot of spearman correlations (≥ .40) between predictors and criterion.

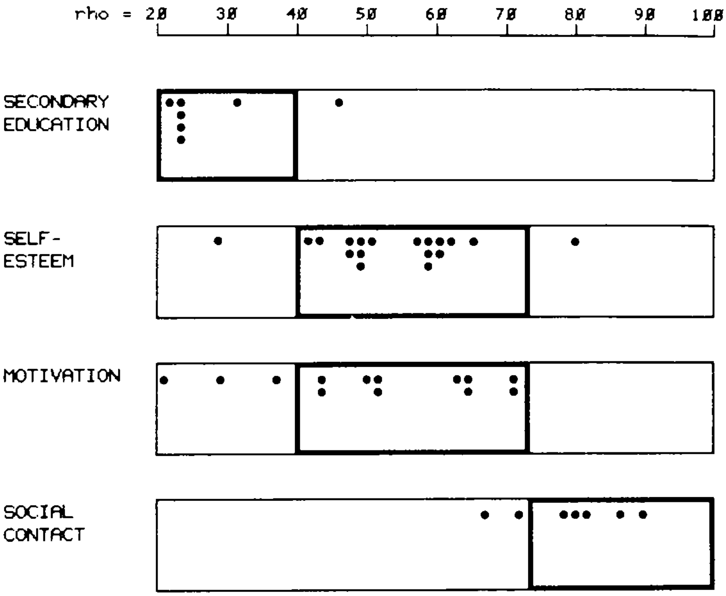


Figure 6

Spearman correlations of predictors with criterion. The predictors are listed according to their position in the weak continguity regions.

4.4. TEST FOR RADEX STRUCTURE

Table 1 shows the alienation coefficients of the MINISSA and CMDA configurations. Imposing regional constraints for our hypothetical radex-structure is payed off by a little increase of the alienation coefficient.

However, to test whether the final CMDA-plot is equivalent with the initial MINISSA plot, Lingoes-Woodroofe correlations, based on the ranks of the correlations/distances are used, instead of the alienation coefficients.

In their first approach, Lingoes and Borg (1980) adapted Hotelling's t-test in order to test whether the difference of the correlation between data and MINISSA-plot on the one side, and the correlation between data and CMDA-plot on the other side, is large/small enough to reject/accept the equivalence between both configurations. According to this stochastic model, we

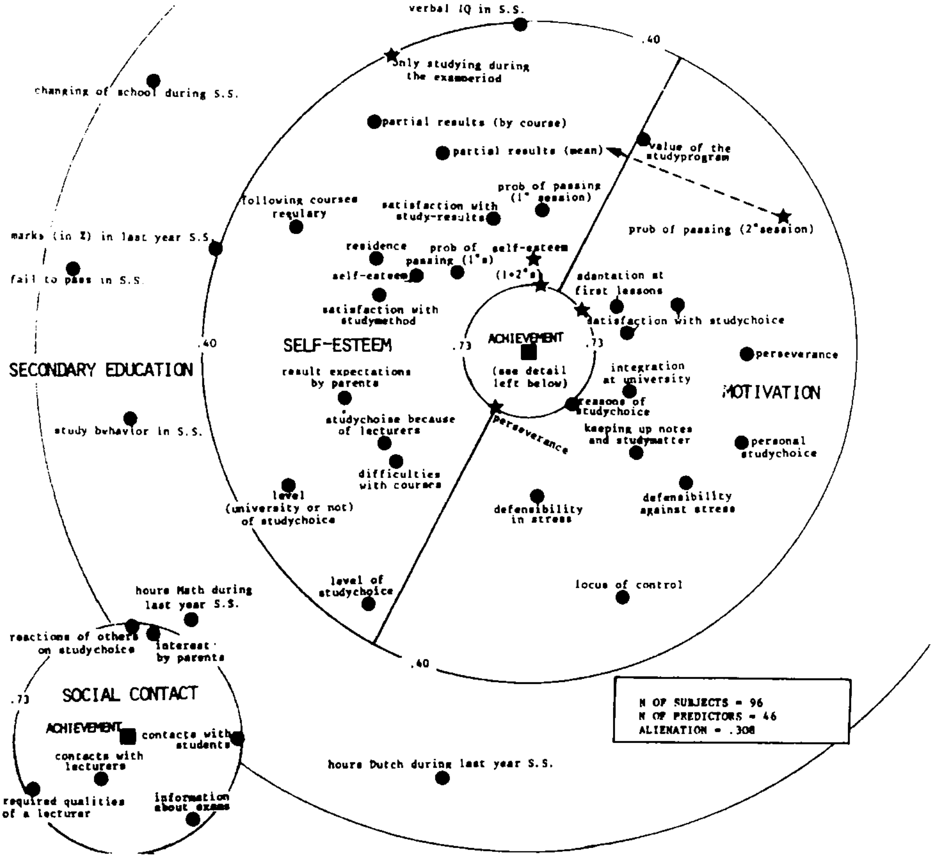

verbal IQ in S.S.
.40
Only studying during the examperiod
changing of school during S.S.
partial results (by course)
partial results (mean)
value of the studyprogram
prob of passing (1° session)
following courses regulary
satisfaction with study-results
prob of passing (2°session)
marks (in %) in last year S.S.
residence
prob of passing (1°s)
self-esteem
self-esteem (1+2°s)
fail to pass in S.S.
adaptation at first lessons
satisfaction with studymethod
satisfaction with studychoice
.40
SELF-ESTEEM
.73
ACHIEVEMENT
(see detail left below)
.73
perseverance
SECONDARY EDUCATION
result expectations by parents
integration at university
MOTIVATION
reasons of studychoice
studychoise because of lecturers
study behavior in S.S.
Perseverance
keeping up notes and studymatter
personal studychoice
difficulties with courses
level (university or not) of studychoice
defensibility in stress
defensibility against stress
level of studychoice
locus of control
hours Math during last year S.S.
reactions of others on studychoice
interest by parents
.40
.73
SOCIAL CONTACT
N OF SUBJECTS = 96
N OF PREDICTORS = 46
ALIENATION = .308
ACHIEVEMENT
contacts with students
contacts with lecturers
hours Dutch during last year S.S.
required qualities of a lecturer
information about exams

Figure 7

CMDA

Spatial plot of spearman correlations (α ≤ .05) with regional constraints.

must reject the CMDA configuration (see Table 2).

alienation	MINISSA (no constraints)	.294
	CMDA (with constraints)	.308
increase of alienation		.014

Table 1

Guttman-Lingoes alienation coefficients of initial (MINISSA) and final (CMDA) configuration.

Lingoes-Woodroofe correlation between	
data - MINISSA: rho_1	.783
data - CMDA : rho_2	.767
Probability of difference between rho_1 and rho_2	.001

Table 2

Lingoes-Borg adaption of Hotelling's t-test for testing the difference between Lingoes-Woodroofe correlations.

Later on, this approach was criticized by Lingoes and Borg (1982), based on the arguments that the underlying models from which it was adapted "fail to be satisfactory, in particular because their error models involve unacceptable assumptions for MDS applications".

Their second approach (Lingoes and Borg (1983a)) consists of quasi-statistical decision model, in which several decision levels or rules have to be satisfied in order to accept the equivalence. This decision-model is based on the partial correlation between MINISSA and CMDA, rho_c, which must

1) share more variance than it doesn't, i.e. exceed $\sqrt{.5}$
2) exceed 3 times the degradation coefficient K, where K is the proportion of variance not explained by the raw correlation between MINISSA and CMDA ($K = \sqrt{1-rho_3^2}$).

Based on this new approach, the CMDA configuration in our example (table 3) was accepted as equivalent with the initial MINISSA configuration.

Very recently, Lingoes and Borg (1983b), developed a new mathematical decision model (GUIDER), that replaces rather arbitrarily rule 3 of the previous model. The basic idea is that at best 7% of all possible final configurations, generated by their GUIDER model, should be closer to the initial

MINISSA-configuration than the actual observed final configuration of the CMDA-program. We agree with Lingoes and Borg that this "present proposal has much to recommend it in terms of being less arbitrary and more informative in the realm of the possible". Our example does comply with this new requirement (see table 4). In other data analysed up till now, it seems that the previous decision rule (Ratio > 3) and the present rule (proportion of Ratio < 7%) reach close results in terms of accepting/rejecting the equivalence hypothesis.

We just have given a very brief overview of recent developments with regard to the crucial question whether one can accept the radex-constraints imposed on the data. It seems to us that the latest proposed GUIDER model gives a conclusive answer to that question, not only in mathematical terms, but also in terms of close agreement with what one should expect from visual inspection of the constraint configuration.

L-W correlation between MINISSA - CMDA : rho_3	.967
partial correlation of rho_3 : rho_c	.919
degradation coefficient $K_3 = \sqrt{1-rho_3^2}$	.254
ratio $R = rho_c / K_3$	3.621
decision-model	
first rule:	
do both plots have more variance in common, or $rho_c > \sqrt{.5}$?	yes
second rule:	
does rho_3 exceed 3 times degradation K_3, or ratio $R > 3$?	yes
decision:	
the equivalence of both configurations is	accepted

Table 3

Lingoes-Borg quasi-statistical model for choosing between two alternative configurations.

P(R > w) = proportion of observed ratio's exceeding all possible ratio's	.044
G.U.B. = greater upper bound of P(R > w), over all possible delta's	.045
decision-model	
first rule:	
do both plots have more variance in common, or $\text{rho}_c > \sqrt{.5}$?	yes
second rule: the GUIDE distribution	
P(R > w) < 7% ?	yes
G.U.B. < 7% ?	yes
decision:	
the equivalence of both configurations is	accepted

Table 4

Lingoes-Borg mathematical decision model for choosing between two alternative configurations.

5. CONCLUSION

When a large number of different variables are to be selected in order to find distinct and predictive predictors for regression-analysis or more complex causal models, our approach, which is in its essence a method to structure the original data according to a radex, can be helpful.

This approach does not have the disadvantage of other "blind" structuring techniques like cluster- or factor-analysis, which are only sensitive to correlations between predictors, and neglects the criterion which has to be predicted. In our radex-approach, we reach two objectives at once: discriminant validity, which means the grouping of interrelated variables, and predictive validity which groups the predictors according to their relation with the criterion.

This approach has also some disadvantages. To find a radex is not so easy, especially when there are more than 2 dimensions to look at. On the other hand, once a radex is found and accepted, the spatial result is more accessible for the readers, to whom the interplay between discrimination and prediction becomes more visible now. Up till now, the largest number of variables treated was 83. This number can, of course, be larger, if the alienation does not increase too much. The boundaries of our approach do not lie in the number of variables, but in the accuracy and interpretability of the spatial representation.

Another, technical, disadvantage is the large number of regional constraints, which is often over 100 for a radex-structure (more complex than the present example) in 70 or more points. These constraints, taken together with both matrices of the original correlations and initial configuration, require a large number of computer-memory or, alternatively, computer-time. A simpler convertion of a radex in regional constraints is welcome.

REFERENCES

[1] Blalock, M., Causal models in the social sciences, Chicago, Aldine (1971).

[2] Borg, I., and Lingoes, J.C., A model and algorithm for multidimensional scaling with external restrictions on the distances, Psychometrika, 45 (1980), 25-38.

[3] Borg, I., and Lingoes, J.C., Multidimensional scaling with side constraints on the distances, In J. Lingoes, E. Roskam, and I. Borg (Eds.), Geometric representations of relational data, Ann Arbor, Mathesis Press, (1979), 733-790.

[4] Denison, D.P., Multidimensional scaling and structural equation modelling: a comparison of multivariate techniques for theory testing, Multivariate Behavioral Research, 17 (1982), 447-470.

[5] Guttman, L., A new approach to factor analysis: the radex, In P.F. Lazersfeld (Ed.), Mathematical thinking in the social sciences, New York, Free Press (1954).

[6] Guttman, L., A general nonmetric technique for finding the smallest coordinate space for a configuration of points, Psychometrika, 33 (1968) 469-506.

[7] Kruskal, J.B., Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, Psychometrika, 29 (1964a), 1-27.

[8] Kruskal, J.B., Nonmetric multidimensional scaling: a numerical method, Psychometrika, 29 (1964b), 115-129.

[9] Lingoes, J.C., The Guttman-Lingoes nonmetric program series, Mathesis Press, Ann Arbor, Michigan (1973).

[10] Lingoes, J.C., Testing regional hypotheses in multidimensional scaling, In Borg, I. (Ed.) Multidimensional data representations: when and why, Mathesis Press, Ann Arbor (1981).

[11] Lingoes, J.C., and Borg, I., CMDA-U: confirmatory monotone distance analysis - Unconditional, Journal of Marketing Research, 15 (1978), 610-611.

[12] Lingoes, J.C., and Borg, I., A model for choosing between alternative configurations derived from ordinally constrained data, University of Michigan, Computer Centre, technical report, 1 (1981), 1-28.

[13] Lingoes, J.C., and Borg, I., A critique of two proposals for comparing MDS configurations, University of Michigan, Computing Centre, technical report, 1 (1982), 1-26.

[14] Lingoes, J.C., and Borg, I., A quasi-statistical model for choosing between alternative configurations derived from ordinally constrained data, British journal of mathematical and statistical psychology, 36 (1983a).

[15] Lingoes, J.C., and Borg, I., A mathematical decision model for the Lingoes-Borg criteria in testing the equivalence hypothesis. University of Michigan, Computer Centre, technical report, 1 (1983b), 1-20.

[16] Lingoes, J.C., and Roskam, E., A mathematical and empirical study of two multidimensional scaling algorithms, Psychometric Monographs, 38 (1973).

[17] Shepard, R.N., The analysis of proximities: multidimensional scaling with an unknown distance function, I, Psychometrika, 27 (1962a), 125-140.

[18] Shepard, R.N., The analysis of proximities: multidimensional scaling with an unknown distance function, II, Psychometrika, 27 (1962b), 219-246.

[19] Van Overwalle, F., Region-program (Fortran 4) to compute ordinally constraints for testing regions in the Lingoes-Borg CMDA-program (1983).

I would like to thank Prof. Dr. J. Lingoes for his articles on recent developments in equivalence testing and for his comments on this paper. I am also grateful to B. Mergaerts and M. Van der Post who helped with the English translation.

ADDENDUM

One can make objections against the rather unusual level of 7% in the Lingoes-Borg (1983b) guider-decision model to accept or reject the equivalence of configurations with or without structural constraints. Fortunately, another heuristic mdethod was developed by Hubert and Golledge (Hubert, L.J. and Golledge, R.G., A heuristic method for the comparison of related structures, Journal of Mathematical Psychology, 23 (1981), 214-226). Similar in its possible applications, their heuristic can also be used to test the equivalence of configurations derived from the same data.

The Lingoes-Borg and Hubert-Golledge approaches have in common that they are not based on a theoretical distribution, but on a reference distribution instead. There are however, important differences between both approaches, namely in the generation of the reference distribution. Lingoes and Borg (1983b) generate a reference distribution of partial correlations, and hence derived ratios. This reference distribution is limited by the two raw correlations between both configurations and the data, increasing from zero to their actual values. If less than 7% possible ratios of this reference distribution exceeds the actual ratio, the equivalence can be accepted. Hubert and Golledge, on the other hand, permutate randomly the data matrix of (dis)-similarities and compute for each permutation a correlation coefficient. If the observed correlation does not differ significantly (e.g. at a level of .01 or .05) from all correlations generated by these random permutations, they contend that the difference between both configurations is not significant.

In the next table, we give some numerical examples. In our examples (67 or more variables), a firm similarity between the Lingoes-Borg and Hubert-Golledge results appears. One can observe that Lingoes-Borg's 7% lies somewhat half-way between the probability level of 1% and 5% of the Hubert-Golledge heuristic. However, in Lingoes' examples, which consist of 8 and 17 points, both methods do not fit. The Lingoes-Borg guider is much more difficult to satisfy than the Hubert-Golledge heuristic. In order to accept these smaller configurations, Lingoes and Borg (1983b) refer to their previous approach (Lingoes and Borg (1983a)), which is even more arbitrary. To avoid this, one can use the Hubert and Golledge (1981) heuristic model.

DESCRIPTION OF DATA			LINGOES-BORG GUIDE		HUBERT-GOLLEDGE HEURISTIC			
data	hypothesis	variables	ratio	proportion of ratio	permutations	r_s	z-value	probability of difference
GLUSHKO	EQUIVALENCE	17	1.004	.328	1000	.129	1.219	.107**
GUTTMAN	DOUBLE-FACET	8	1.123	.295	800	-.101	-.223	.430**
INTEGRATIE(5)	FINAL (1)	75	3.299	.052*	1000	-.097	-1.764	.038*
INTEGRATIE(5)	WEINER 1	75	2.808	.070*	1000	-.092	-1.883	.034*
INTEGRATIE(5)	WEINER 2	75	3.834	.040*	1000	-.062	-1.082	.125**
BLOK (2)	FINAL (3)	67	3.031	.061*	1000	-.105	-1.758	.035*
BLOK (2)	WEINER 1	67	3.065	.059*	1000	-.079	-1.298	.089**
BLOK (2)	WEINER 2	67	3.204	.054*	1000	-.075	-1.149	.135**
PMS (2)	FINAL (5)	81	2.418	.087	1000	-.122	-2.521	.007
			* equivalence accepted (proportion < .07)		** equivalence accepted (probability of difference > .05) * equivalence accepted (probability of difference > .03)			

Table 5

Thanks to Dr. G. De Soete for bringing the possibilities of the Hubert and Golledge heuristic to my attention.

This Page Intentionally Left Blank

AUTHOR INDEX

Abdi, H. 6,7
Adams, E.W. 286
Andersen, E.B. 168,257,414,422
Anderson, N.A. 37
Andrich, D. 299,414,422,429
Anthonisse, J.M. 366
Arabie, P. 5,223,379
Arbuckle, J. 240
Arend, L.E. 67
Armstrong, W.E. 184
Arrow, K. 180,365,374
Atkinson, R.C. 122,124,125,135
Ayer, M. 264

Bagozzi, R.P. 318
Bailey, D.E. 378,379
Balinski, M.L. 366
Barlow, R.E. 241-244,246,249,264, 266-268
Barthélémy, J.-P. 183,184
Bartholomew, D.J. 241,242,244,246
Bartlett, F.C. 3,14,17
Bass, F.M. 387
Bauer, H. 319,341
Beals, R. 194,227,359
Benzecri, J.P. 7
Berg, J.A. 417
Berge, C. 219
Betson 126
Birkhoff, G. 230
Birnbaum, A. 167
Birnbaum, M.H. 37,303
Blake, R. 60
Bock, R.D. 257
Bonacich, P. 219
Boorman, S.A. 5,223
Borg, I. 360,453,460,463,465-467, 470
Bouchet, A. 274
Bouglé, C. 219
Bousfield, W.A. 7
Bower, G.H. 122,124
Bradley 427
Breimann, L. 319,341
Bremmer, J.M. 241
Brien 125
Broadbent, D.E. 89
Bruner, J. 123
Brunk, H.D. 241
Buehler, J.N. 67
Buffart, H. 42-44,51,56,57
Buneman, P. 8
Bush, R. 122,124,150

Campbell, D.T. 318
Carroll, J.D. 3,8,380-382,384-386, 389,390
Chang, J.J. 3,8,380
Chapanis, A. 127,129
Chipman, J.S. 275
Clark, L.A. 380,385,389,390
Cogan, A.I. 60
Cogis, O. 274
Cohen, M.A. 60,67,79,84
Condorcet 183,185
Conway, J. 61
Cook, T.D. 318
Coombs, C.H. 159,171,283,383,415,438
Cornsweet, J.C. 60
Cornsweet, T.N. 60
Cowan, T.M. 105
Cozzens, M.B. 280
Crothers, E.G. 124
Cunningham, J.P. 3,8,11,247,380

Dawes, R.M. 438
Degenne, A. 7,219,224
de Haan, B. 60
Denison, D.R. 450
DeSarbo, W.S. 378-380,384,385,387, 389,390
De Soete, G. 380,384
de Weert, C.M.M. 61,74
Dobson, A.G. 386
Dobson, J. 8
Doignon, J.-P. 210,216,231,233,234, 274,359
Domhoff, W. 219
Doreian, P. 219
Douglas, G.A. 296
Draper, S.W. 105
Drösler, J. 230,359
Ducamp, A. 210,215,216,231,233,234, 274,293,297,300
Duval, S. 177
Duval, V.H. 177

Ebbinghaus, H. 121,125
Ehrenstein 70,72-75
Engelman, L. 379

Escher, M.C. 105
Estes, W.K. 122

Falmagne, J.-C. 210,215,216,231, 233,234,256,274,286,293,297,300, 359
Farris, J.S. 385
Fischer, G.H. 160,296,298,393,398, 399,420-422
Fishburn, P. 193,194,210,211,216, 272,278,360
Flament, C. 7
Fox, R. 60
Franke, C.H. 285,286
Friendly, M. 3,6,27,28
Furnas, G.W. 380-382,384,385

Galanter, E.H. 171,284
Gänsler 319,341
Garner, W.R. 33,56
Gati, I. 178
Geissler, H.-G. 34,35,37,42,43,56
Gellatly, A.R.H. 61
Gerrits, H.J.M. 60,80
Gill, P.E. 263
Golledge, R.G. 470,471
Goodman, N. 284
Gordon, A.G. 184
Green, D.M. 87,161
Greeno, J.G. 122
Gregory, R.L. 105,106
Grossberg, S. 60-63,67,70,73,75,78, 79,84
Guilbaud, G.Th. 183
Guilford, J.P. 258,259,265,267
Gulliksen, H. 398
Gustaffson, J.E. 399
Guttman, L. 210,231,274,293,297, 300,414,415,452,454,455,465
Gvozdev, Y.A. 152

Hakami, S.L. 8
Halmos, P.R. 319
Hamilton, A. 368,369,373
Hanter 127
Harbeck, A. 61
Harshman, R. 390
Harter 125
Hartigan, J.A. 8,244,247,379,380, 385
Hasse 186,187,229
Hayes-Roth, B. 56
Hayes-Roth, F. 56
Hebb, D.O. 60
Heise, D.P. 318
Hendrickson, A.E. 66
Henley, N.M. 3,6,7,14,25,26
Herbst, K. 230
Heron, W. 60
Hick, W.E. 34
Hildreth, E. 62
Hill, M.O. 7
Hirotsu, C. 246
Hoffmann, J. 40
Hohle, H. 259
Holman, E.W. 350
Hoogstraten, J. 147
Horton, J.C. 66
Hotelling 453,463,465
Hubel, D.H. 66
Hubert, L.J. 470,471
Huffman, D.A. 105,113
Hull, C.L. 122
Hunt, S.P. 66

Jacquet-Lagrèze, E. 280
Jaffray, J.Y. 359
James, G.S. 246
Jansen, M.G.H. 409
Jansen, P.G.W. 420,424
Jastrow, J. 160,166
Jerison, H.J. 89,91,92
Johnson, S.C. 6,385
Jones, L.V. 257
Joreskog, K.G. 394
Julesz, B. 61

Kanizsa, G. 61
Kaufman, L. 60
Kempf, W.F. 296
Kendall, M.G. 184
Kennedy, J.M. 61,78
Kerlinger, F.N. 6
Kling, G.W. 122
Klix, F. 37
Kow 126
Krantz, D.H. 194,197,204,227,239, 256,259,295,348,349,359,360
Krinchik, E.P. 136
Kristof, W. 393,398
Kristofferson, A.B. 89
Kruskal, J.B. 7,12,248,265,452

Laming, D. 93,99
Lanczos, C. 172
Land, E.H. 60,77,78
Larimer, J. 240
Lawley, D.N. 399
Leeuwenberg, E.L.J. 42-44,56,57, 61
Lehmann, D.R. 386
Leonard, T. 407
Leontyev, A.N. 136
Levey, J. 61
Levitt, P.R. 223
Lewis 409
Likert, R. 414,429
Lindley, D.V. 405,407
Lingoes, J.C. 415,452,453,455-457, 460,463,465-467,470
Livingstone, M.S. 66
Lockhead, G.R. 67
Loève, M. 319,341
Loh, H.K. 160
Lomov, B.F. 127,128,155
Lord, F.M. 393,399,405,406,420
Lorrain, F. 221,223
Luce, R.D. 88,122,209,239,240,256, 271,272,277,279,284,286,299,301, 348,359,427
Luong, X. 10

Mac Donald, R.R. 241
Mac Rae, A.W. 160
Magazanik, Y.M. 147
Mandel, J. 241
Marks, W. 61
Marr, D. 62
Masters, G.N. 414,429
Mausfeld, R. 257
Mac Clelland, G. 240
Mc Cornick, W.T. 379
Meyer, F. 101
Miller, G.A. 3,6,7,22,23,378
Mingolla, E. 60,62,70,75,78
Monjardet, B. 183,184,210,211,274
Mosteller, F. 122,124,150,259
Murray, W. 263

Narens, L. 286,359
Necker, L.A. 106
Ng, Y.-K. 193
Nicely, P.E. 378
Nishisato, S. 7
Novick, M.R. 393,398,399,405,406

O'Brien, V. 59,67,69,71
Orth, B. 359
Owen, J. 405,407

Parks, T.E. 61
Pavlov 121
Pendl, P. 296
Penrose, L.S. 105,106,113
Penrose, R. 105,106,113
Pessimier, E.A. 387
Petry, S. 61
Pfanzagl, J. 286
Pickett, R.M. 89,91,92
Platonov, K.K. 127
Poincaré, H. 284
Popper, K.R. 394
Prenowitz, W. 115
Pritchard, R.M. 60
Pronco, N. 155
Pruzansky, S. 12,380-382,385,386,389
Puffe, M. 34,35,37,42,43,56

Rainio, K. 438
Ramsey, J.O. 259,262,295
Rao, S.S. 384
Rapaport, G.M. 417
Rasch, G. 257,293-297,299-301,304, 305,393,399-401,405-407,413,414, 419-423,426,427,429
Ratliff, F. 60
Redies, C. 61,70,73
Restle, F. 42,57,284
Riblier, W. 221
Riggs, L.A. 60
Riguet, J. 210,274
Riordan, J. 8
Roberts, F.S. 211,271,272,275,276, 280,283-286,288,289,349,353
Robinson, R.E. 286
Rosch, E. 40
Rosenberg, S. 3,5
Roskam, E.E. 265,294,296,299,300, 360,414,420,452
Roubens, M. 214,215,272,280
Roy, B. 214,215,280
Russo, J.E. 256
Rybalchenko, M.V. 147

Samejima, F. 414,429
Sarle, W.S. 379
Sarler, L.D. 221

Sattath, S. 3,8,9,11,380,381
Schaber, G. 394
Scheffe, H. 248
Scheiblechner, H. 296
Scheidereiter, U. 37
Schiffman, S. 7
Schönemann, P.H. 360
Schwarz, L.M. 127
Schweitzer, P.J. 379
Scot, D. 210,257,284
Seidman, S.B. 219
Shepard, R.N. 379,452
Shorack, G.R. 243,245-247
Simon, H.A. 321
Smith, A.F.M. 405
Smith, P.T. 241
Sneath, P.H. 8
Sneider, F. 155
Sokal, R.R. 8
Solomon-Winn 124
Spada, H. 296
Spillman, L. 61,70,73
Stegmüller, W. 322
Stenson, H.H. 89,91,92
Stern, W. 37
Stevens, S.S. 171
Steyer, R. 317,318,321,323,330,
331,340,341
Stone, M.H. 407
Stumpf, H. 257
Stute, H. 319,341
Suck, R. 359
Sugihara, K. 105
Suppes, P. 210,239,256,284-286,
296,301,322,348
Swets, J.A. 87,89,161
Swift 125

Tergan, S.O. 100
Térouanne, E. 105,106
Terry 427
Thompsen 205,206
Thorndike, E.L. 121
Thurstone, L.L. 257,258
Timmerman, J.G.M.E.N. 60
Torgerson, W.S. 7,159
Tryon, R.C. 378,379
Tukey, J.W. 239-241
Tulving, E. 7
Tversky, A. 3,8,9,11,12,177-180,
182,188,193,194,227,239,256,
348,359,360,380,381,438

Vallacher, R.R. 5
Van Acker, P. 194
Van der Ven, A.H.G.S. 393,405,407
van Eeden, C. 264
Van Heerden, J. 417
Van Overwalle, F. 457,460
van Tuijl, H.F.J.M. 61,74
Venda, V.F. 130,131,136,137
Vendrick, A.J.H. 60,80
Verges, P. 7
Vincent, S.B. 133,134
Vincke, Ph. 214,215,272,280

Wegner, D.M. 5
White, H.C. 221,223
White, T.W. 379
Wiesel, T.N. 66
Winer, B.J. 243,245,249
Winsberg, S. 259,262
Wold, H. 386
Woodroofe 463,465
Woodworth 122,125,127
Wright, B.D. 407
Wu, J.-Y. 66
Wykoff, L.B. 90

Yannakakis, M. 234
Yarbus, A.L. 60,75
Yan, S.S. 8
Young, F.W. 7
Young, H.P. 366

Zarakovsky 130,134
Zimolong, B. 101
Zinnes, J. 284-286

SUBJECT INDEX

adaptation dynamics 121
adaptive 60,61
additive 257,263,295,347-353,357,358,377,378,380,385,386,389,390
additive difference 193-195,197,202,205
additivity 239,241,244,247-249,334-336
asymmetry 177-180,182-184,186-188
attention 87,89-93,98,100-102
attitude theory 294,304,305

binocular rivalry 59,60,63,64,79,80
biorder 210-212,214-216,227,232,235,236,274
biorder dimension 227,231,233-236
bipartition 4,5,12,14
brightness perception 59-62,67

categories 33,36,41-46,48,51,52,54,55
causal model 317-319,321,330,333-338,340,451,468
choice 34,35,56,88,94,101,159,160,172,255,299
chromatic number 234
classification 33-35,37,38,40-42,45-50,52-55
clustering 3,6,7,13,14,378-380,382,385,387-389,393,394,396,397,450,452,467
coding 33,43-45,52,54,55
complementary color induction 59
complete order 183,185,187
composite measurement 293,300
condorcean analysis 180,183-186
Condorcet relation 183,185
conjoint measurement 205,239,241,244,257,347,348,350
contour perception 59
convex 106,110,111,113,116,264,457
Craik-O'Brien effect 59,67,69,71

detectability 255
detection 87-102,161
dichotomization 393,395,413,414,416-419,422-428
dichotomous 297,332
discrimination 87,88,90,91,93,95-97,100,101,169-172,255,257,303

Effet Condorcet 183
equiorderedness 319,321,324-330,332,333,336

Ferrers relation 210,272-275
filling-in 35,59-61,64,67,68,75-79,84,85
form perception 59-62
fuzzy order 309

game theory 433,435,438,439,441
graph 7,111,113,116,219,220,222-224,234,283,284,286,289,310,314
guided inference 32-34,41-43,54-56
Guttman scale 210,293,297

homogeneous family 271,272,275,276,279,280
hypergraph 219,220,223,224,234

illusory contours 59
illusory figure 61,62,70,78
impossible figure 105,106,108
incomparability 309,312,314
indifference 209,210,214,309,312,314,351-353,356,358,359
information 33
interpretation 105,106,111,113,115,116,118
interval order 209-211,216,274,309-311,313-314
interval scale 347
intransitivity 193,194
isotone regression 239,241-244,246-249

join geometry 115

Λ-relation 271-273,276-280
Land retinex theory 59-61,77-78
latent trait theory 159,160,163-165,169,171
lattice 227-230,236
lattice dimension 227-230,236
learning 35,36,91,121-127,129-133,135,137-156,255
learning plateau 124-127,133,137,145,147,155
lexicographic ordering 193,194
linear order 180,183,184,264,265,267,272,273,275,277,278

majority rule 177,180,185
Markov chain 438
Markov process 122,123,138
matching 59,63,64,66,79,80,85,283,284,286-288,319,340
meaningful 227,283-287,303-305,333,435
measurement 56,101,177,257,259,271,283,284,286,293,295,296,303-305,349,353,380,381
median order 177,180,183-188
multidimensional scaling 7,227,348,359,360,383,390,449,451,452

neon color spreading 59,61,70,73
NP-complete 234

observation 87,98
order(ing) 56,108,110,115,116,118,119,169,171,172,239-245,263-267,279,283,284,286,287,289,293,296-300,302-304,357,359

partial interval order 309,312-314
partial order 184,187,228,246,264,265
perception 59-62,67,79,89,91,93,101,102,105,136,139,149,159,171,353
perceptual 37,42,160,161,166,167
perceptual decision 87-90,97,101
polychotomization 413,414,416,418-420,423,427,429
polyhedral 108,110,111,113,115,116,118,119
preference 160,193,194,203,206,209-212,214,215,255,260,268,284,309,312-314,377
preorder 179,309,310,312-314
preorderedness 317,319,321-326,328-336,338-340
prisoner's dilemma 433,435,438,439,442
probabilistic consistency 271,272
proximity (distance) matrix 3,5,7,8,12-17,19-21,25,27,348,377-381,383,384,389,390
pseudo-order 309

quasiorder 273,274,277,349,358
questionnaire 305,414,415,417,427

Radex structure 449,454-457,459,460,463,466-468
Rasch model 257,293,294,296,299-301,304,305,393,400,401,405,407,413,414,419,420,423,424,426,427,429
reaction-time 34,37-40,43,45,48
réseau 219-221,223-224
réseau social 219,220

semi-order 13,193,209-212,271,272,275-277,280,283-286,288,309
similarity 6-8,10,11,39,177-179,181-183,185,186,188,190,193,220,223,227,271,272,283,360,381,396,397,452-454,470
simple order 211,212,246,288-290,298,351,356
specific objectivity 293-296,303-305
stabilized images 59-61,75,79,80
strategy 34,40,121,124,133,136-143,145-147,149-152,155,156
strict order 325

threshold 7,86,88,89,92,96,98,209-212,214-216,257,272,278,280,284
transitivity 115,116,183,185,195,197,215,220,222,256,272,273,283,284,297,298,300,310,313,314,325-327,329,330,349,354,357
tree 3,7,8,10-12,15,17,19,24,26,28,37-40,47,49,50,52,53,55,377,378,380-383,385-387,389,390

unfolding 171,380-383
utility 209-211,256,257,271

visual cortex 59

weak causal model 317-319,338-341
weak order 195,197,198,205,231,272-277,279,328-330,347-350,353,356,358,359,423